

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:49:27 ; Search time 117 Seconds

(without alignments)
5507.858 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcagtgagcagagagagcag.....agagtgacctaccagtg 1460

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database: Issued Patents NA:*

1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*\n2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*\n3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*\n4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*\n5: /cgn2_6/ptodata/2/ina/PCTUS.COMB.seq:*\n6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	92	6.3	804 4 US-08-932-411A-19	Sequence 19, Appl
2	23	1.6	738 4 US-08-723-570-13	Sequence 13, Appl
3	23	1.6	738 4 US-08-932-411A-13	Sequence 13, Appl
4	23	1.6	1333 1 US-08-910-973-21	Sequence 21, Appl
5	23	1.6	1333 4 US-09-499-227-21	Sequence 21, Appl
6	23	1.6	1385 4 US-08-932-411A-17	Sequence 17, Appl
7	21	1.4	310 1 US-08-552-142A-12	Sequence 12, Appl
8	21	1.4	1268 1 US-08-910-973-12	Sequence 12, Appl
9	21	1.4	1268 4 US-09-499-227-12	Sequence 12, Appl
10	21	1.4	1352 1 US-08-552-142A-10	Sequence 10, Appl
11	21	1.4	1535 1 US-08-910-973-10	Sequence 10, Appl
12	21	1.4	1535 4 US-09-499-227-10	Sequence 10, Appl
13	14	1.4	1550 3 US-09-234-332-3	Sequence 3, Appl
14	13	1.3	50 3 US-08-358-627F-4	Sequence 4, Appl
15	13	1.3	50 3 US-08-793-044-11	Sequence 11, Appl
16	13	1.3	50 4 US-08-465-712C-4	Sequence 4, Appl
17	13	1.3	50 4 US-09-552-733-4	Sequence 4, Appl
18	18	1.2	480 1 US-08-438-123-16	Sequence 16, Appl
19	18	1.2	1336 4 US-09-016-434-129	Sequence 129, Appl
20	18	1.2	1462 1 US-08-552-142A-16	Sequence 16, Appl
21	18	1.2	1951 1 US-08-910-973-16	Sequence 16, Appl
22	18	1.2	1951 4 US-09-499-227-16	Sequence 16, Appl
23	18	1.2	3446 4 US-09-620-312D-653	Sequence 653, Appl
24	18	1.2	4066 4 US-09-367-750-1	Sequence 1, Appl
25	18	1.2	4797 4 US-09-419-568F-25	Sequence 25, Appl
26	18	1.2	4797 4 US-09-354-243B-25	Sequence 25, Appl
27	18	1.2	31728 4 US-09-453-702B-64	Sequence 64, Appl

28	17	1.2	24 3 US-08-358-627F-6	Sequence 6, Appl
29	17	1.2	24 4 US-08-465-712C-6	Sequence 6, Appl
30	17	1.2	24 4 US-09-552-733-6	Sequence 6, Appl
31	17	1.2	24 4 US-09-687-731-2	Sequence 2, Appl
32	17	1.2	24 4 US-09-687-731-8	Sequence 8, Appl
33	17	1.2	24 4 US-09-687-731-10	Sequence 10, Appl
34	17	1.2	510 4 US-09-252-991A-9622	Sequence 9622, Ap
35	17	1.2	685 1 US-08-751-782-5	Sequence 5, Appl
36	17	1.2	685 2 US-08-925-171-5	Sequence 5, Appl
37	17	1.2	846 4 US-09-252-991A-9635	Sequence 9635, Ap
38	17	1.2	948 4 US-09-252-991A-14810	Sequence 14810, A
39	17	1.2	954 4 US-09-252-991A-9532	Sequence 9532, Ap
40	17	1.2	1062 4 US-08-086-439C-1	Sequence 1, Appl
41	17	1.2	1161 1 US-08-434-877-1	Sequence 1, Appl
42	17	1.2	1238 4 US-09-183-861-75	Sequence 75, Appl
43	17	1.2	1238 4 US-09-022-765-75	Sequence 75, Appl
44	17	1.2	1238 4 US-09-551-974A-75	Sequence 75, Appl
45	17	1.2	1238 4 US-09-551-974A-75	Sequence 75, Appl

ALIGNMENTS

RESULT 1
US-08-932-411A-19
Sequence 19, Application US/08932411A
Patent No. 6566496
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 804 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 160..801
US-08-932-411A-19

THIS PAGE BLANK (USPTO)

Query Match 6.3%; Score 92; DB 4; Length 804;
Best Local Similarity 100.0%; Pred. No. 9.9e-37;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 762 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 821
DB 463 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGGATGACGCCAACTTACAAAG 522
QY 822 ATGAGACCTTGGCTTCCGCCCACTACAT 853
DB 523 ATCGAGACCTTGGCTTCCGCCCACTACAT 554

RESULT 2

US-08-722-570-13
Sequence 13, Application US/08722570
Patent No. 655537

GENERAL INFORMATION:

APPLICANT: Anderson, David J.

APPLICANT: Ma, Qifu

TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/722,570

FILING DATE: 27-SEP-1996

CLASSIFICATION: 5365

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 738 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

US-08-722-570-13

Query Match 1.6%; Score 23; DB 4; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACACT 754
DB 304 GAGCGCAACCGCATGCACACACT 326

RESULT 3

US-08-932-411A-13

Sequence 13, Application US/08932411A

Patent No. 6566496

GENERAL INFORMATION:

APPLICANT: Anderson, David J.

APPLICANT: Ma, Qifu

TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 31

CORRESPONDENCE ADDRESS:

ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/932,411A

FILING DATE: 15-SEP-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/772,009

FILING DATE: 19-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/722,570

FILING DATE: 19-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 738 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..732

US-08-932-411A-13

Query Match 1.6%; Score 23; DB 4; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACACT 754
DB 304 GAGCGCAACCGCATGCACACACT 326

RESULT 4

US-08-910-973-21

Sequence 21, Application US/08910973

Patent No. 5795723

GENERAL INFORMATION:

APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.

TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectode:

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen O'Connor Johnson KindnessPLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 15:49:38 ; Search time 3511 Seconds
(without alignments)
2493.495 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPQPSGAPVQVTRFERS.....LGATSSACLSFGSLAFSDPL 214

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2888711 segs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters: -DEV-xlh
-MODEL=frame_plus_p2n.model -DB=US0955947/runat_02022004_154933_8046/app_query.fasta_1.391
-O=/cgn2.1/USPTO/US0955947/runat_02022004_154933_8046/app_query.fasta_1.391
-DB=genemb1 -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOFCU=0 -LOOEXT=0
-UNITS=bits -START=-1 -END=-1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=PCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US0955947@cgn.1.1.3508@runat_02022004_154933_8046 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmb1.*
1: gb_da.*
2: gb_htg.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_ov.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sfs.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_da.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_om.*
21: em_or.*
22: em_ov.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sfs.*
28: em_un.*

29: em_vl.*
30: em_htg_hum.*
31: em_htg_inv.*
32: em_htg_other.*
33: em_htg_mus.*
34: em_htg_pln.*
35: em_htg_rtd.*
36: em_htg_mam.*
37: em_htg_vrt.*
38: em_ey.*
39: em_htgo_hum.*
40: em_htgo_mus.*
41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1127	100.0	1330	9	HSAL133776
2	1105	98.0	5340	9	AF234829
3	1099	97.5	165110	9	AL450311
4	1089	96.6	173341	2	AC021954
5	849	75.3	1491	6	A91167
6	849	75.3	1491	6	BD023626
7	849	75.3	1491	10	RNRELAXT
8	849	75.3	258815	2	AC127817
9	819	72.7	861	6	AX698801
10	819	72.7	861	10	MMU76208
11	819	72.7	1861	10	AF364300
12	819	72.7	5567	10	MMU76208
13	819	72.7	138070	2	AC127417
14	819	72.7	215050	2	AC127417
15	523	46.4	170896	5	AC011010
16	388.5	34.5	790	5	GGA012659
17	388.5	34.5	1880	5	AF303000
18	385.5	34.2	1074	5	GGA012660
19	382	33.9	1299	5	AF123883
20	381.5	33.9	770	5	AF123884
21	380.5	33.8	1341	5	AF109014
22	378.5	33.6	1268	6	AR023709
23	378.5	33.6	1268	6	AR225842
24	378.5	33.6	1675	9	BC008687
25	378.5	33.6	1717	9	BC028226
26	378.5	33.6	79970	2	AC011428
27	378.5	33.6	134506	9	AC005738
28	376.5	33.4	1268	9	HSU63842
29	375.5	33.3	1385	10	MMU76207
30	375.5	33.3	10393	10	AF303001
31	374	33.2	258118	2	AC112007
32	374	33.2	258880	2	AC095430
33	370.5	32.9	1412	10	MMU76208
34	370.5	32.9	151044	2	AC011432
35	370	32.8	1527	6	AR308547
36	370	32.8	1527	10	RNU67777
37	368.5	32.7	123855	2	AC102600
38	365	32.4	932	10	MMU76208
39	365	32.4	1315	10	MMU63841
40	365	32.4	1333	6	AR023715
41	365	32.4	1333	6	AR225848
42	365	32.4	149268	2	AC124395
43	362	32.1	26298	2	AC111702
44	359.5	31.9	735	10	MMU67776
45	359.5	31.9	738	6	AR308548

RESULT 1

ALIGNMENTS

HSA133776
 LOCUS HSA133776 1330 bp DNA linear PRI 19-JUN-1999
 DEFINITION Homo sapiens gene for neurogenin 3.
 ACCESSION AJ133776
 VERSION AJ133776.1 GI:5123782
 KEYWORDS bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1
 AUTHORS Ravassard, P., Icard-Hiepkals, C., Wiard, L., Julien, J.P. and Mallet, J.
 TITLE The human neurogenin 3 homolog maps to chromosome 10q21.3 and its expression pattern is identical to that of its murine counterparts
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1330)
 AUTHORS Ravassard, P.
 TITLE Direct Submission
 JOURNAL Submitted (16-MAR-1999) Ravassard P., Igm, CNRS UM69 923, Hopital de la Pitie Salpêtrière, Bat. CERVI, 83 Bd. de l'Hopital, 75013 PARIS, FRANCE
 FEATURES
 source
 1. .1330
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 1. .1330
 /gene="ngn3"
 join(<1.,157,321.,>1330)
 /gene="ngn3"
 1. .157
 /gene="ngn3"
 /number=1
 158. .320
 /gene="ngn3"
 /number=1
 321. .1330
 /gene="ngn3"
 /number=2
 322. .966
 /gene="ngn3"
 /function="early neurogenesis"
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="CAB45384.1"
 /db_xref="GI:5123783"
 /translation="MTPOPSGAPVQVTRTERSPRASEDEVTCPTSAPEPTPTG
 NCABEBCGCGAPRLARRGSRSPKSELALSKORSRKRKANDRENRNDNSA
 IDALRGVLPDPDDAKTETLRPAHNYIMALTOTLRADHSLVALPEPPAPGELG
 SPGGPDMGSLYSPSQSGSLSPASLEERPGILGATSSACLSRGSIAFDPFL"
 BASE COUNT 230 a 459 c 413 g 228 t
 ORIGIN
 Alignment Scores:
 Pred. No.: 8.59e-47 Length: 1330
 Score: 1127.00 Matches: 214
 Percent Similarity: 100.00% Conservative: 0
 Best Local Similarity: 100.00% Mismatches: 0
 Query Match: 100.00% Indels: 0
 DB: 9 Gaps: 0
 US-09-595-947e-10 (1-214) x HSA133776 (1-1330)
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 322 ATGACGGCTCAACCTCGGGTGGCGCCACTGTCCAAGTACCCCGTGAGCGAGCGGTCC 381
 QY 21 PheProArgAlaSerGluAspGluValThrCy6ProThrSerAlaProProSerProThr 40
 DB 382 TTCGCCAGAGCTCGGAGACGAAAGTGAAGTCCCGCCCGCCCGCCAGCCCGCCACT 441
 QY 41 ArgThrProGlyAsnGlyAlaGluAlaGluGluGlyGlyArgArgGlyAlaProArgGly 60

DB 442 CCACACCGGGGAATTCGCCAGAGCGGAGAGGAGGAGCTGCCAGAGGGCCCCAGAGAG 501
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProIysSerGluLeuAlaLeuSerIysGln 80
 DB 502 CTCGGGGGACGGCGCGGGGGACGCGAGCCGGCTTAAAGACGAGTGTGGCTGAGCAAGCAG 561
 QY 81 ArgArgSerArgArgGlyGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 DB 562 CGACGAGTGGCGAAGAGGACCAACACGCGAGCGCAATCGAATGACACGACTCAAC 621
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaIysLeuThr 120
 DB 622 TCGGCACTGAGAGCGCTCGGGGTGTCTGTGCCACTTCCACAGACGAGCAAGCTCAC 681
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIeTPAlaLeuThrGlnThrLeuArg 140
 DB 682 AAGATCGAGACCTCTCGCTTCCGCCCAACTACCTGCGGCTGACTCAACAGCTGCGC 741
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluIupProAlaProHisCy6GlyIleuGly 160
 DB 742 ATAGCGAGCACAGCTTGACGCTGAGCGCGCGCGGCGGCGGAGGAGCTGGGCG 801
 QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTySerProValSerGlnAlaGly 180
 DB 802 AGCCAGAGCGGCTCCCCCGGGGACTGGGGGTCTCTACTCCCACTCCAGGCTGGC 861
 QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuGlyAlaIleThrSerSer 200
 DB 862 AGCTGAGTCCCGCGCTGCTGAGAGGAGGACCGGGCTGTGGGGCCACTCTTCC 921
 QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 922 GCTGCTTGAGCCAGGAGTCTGCTTCTCAGATTTCG 963
 RESULT 2
 AF234829 5340 bp DNA linear PRI 19-OCT-2001
 LOCUS AF234829
 DEFINITION Homo sapiens neurogenin 3 gene, complete cds.
 ACCESSION AF234829
 VERSION AF234829.1 GI:13183002
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 5340)
 AUTHORS Iwaseki, N., Ogata, M., Iwamoto, Y., German, M.S. and Bell, G.I.
 TITLE Mutations in the coding region of the neurogenin 3 gene (NEUROG3) are not a common cause of maturity-onset diabetes of the young in Japanese subjects
 JOURNAL Diabetes 50 (3), 694-696 (2001)
 MEDLINE 21140923
 PUBMED 11246894
 REFERENCE 2 (bases 1 to 5340)
 AUTHORS Lin, J. and German, M.
 TITLE Direct Submission
 JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of California San Francisco; 513 Parnassus Ave., San Francisco, CA 94143-0534, USA
 FEATURES
 source
 1. .5340
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 <3022.,>3666
 /product="neurogenin 3"
 /codon_start=1
 3022. .3666
 /codon_start=1
 /product="neurogenin 3"
 /protein_id="AAK15022.1"
 /db_xref="GI:13183003"


```
repeat_region /note="AluSq repeat: matches 1. .301 of consensus"
14577. .14633
/note="AluSq repeat: matches 243. .299 of consensus"
repeat_region 14843. .14943
/note="MER86 repeat: matches 9. .109 of consensus"
repeat_region 15404. .15613
/note="MIR repeat: matches 3. .213 of consensus"
repeat_region 16626. .16705
/note="MIR repeat: matches 60. .140 of consensus"
repeat_region 16928. .16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region 17476. .17559
/note="MER81 repeat: matches 2. .114 of consensus"
repeat_region 17719. .18069
/note="L1MC4 repeat: matches 7617. .7977 of consensus"
repeat_region 18117. .18345
/note="MIR repeat: matches 7. .262 of consensus"
repeat_region 19169. .19400
/note="MIR repeat: matches 5. .239 of consensus"
repeat_region 19436. .19627
/note="L1MC5 repeat: matches 7728. .7917 of consensus"
repeat_region 19642. .19716
/note="MIR repeat: matches 48. .131 of consensus"
repeat_region 19969. .20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region 20623. .20704
/note="41 copies 2 mer gt 85% conserved"
repeat_region 20738. .20773
/note="U2 repeat: matches 1. .36 of consensus"
repeat_region 20802. .20853
/note="MLT1J repeat: matches 1. .62 of consensus"
repeat_region 20889. .21263
/note="MLT1F repeat: matches 188. .541 of consensus"
repeat_region 21463. .21618
/note="MIR repeat: matches 46. .192 of consensus"
repeat_region 22019. .22326
/note="AluDb repeat: matches 1. .306 of consensus"
repeat_region 22381. .22564
/note="U2 repeat: matches 2453. .2629 of consensus"
repeat_region 22896. .23174
/note="MLT1J repeat: matches 117. .413 of consensus"
repeat_region 23215. .23346
/note="MIR repeat: matches 48. .188 of consensus"
repeat_region 23388. .23532
/note="U2 repeat: matches 2097. .2230 of consensus"
repeat_region 23533. .23837
/note="AluDb repeat: matches 1. .303 of consensus"
repeat_region 23838. .24137
/note="U2 repeat: matches 1754. .2097 of consensus"
repeat_region 24291. .24581
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 24653. .24850
/note="MIR repeat: matches 1. .200 of consensus"
misc_feature 26189. .26685
/note="Cpg Island"
/evidence="not_experimental"
misc_feature 29453. .30918
/note="Cpg Island"
/evidence="not_experimental"
repeat_region 31084. .31156
/note="MIR repeat: matches 65. .138 of consensus"
repeat_region 31618. .31876
/note="AluDb repeat: matches 29. .275 of consensus"
repeat_region 32767. .32830
/note="MIR repeat: matches 76. .139 of consensus"
repeat_region 33050. .33178
/note="43 copies 3 mer tcc 72% conserved"
misc_feature 35112. .36201
/note="Cpg Island"
/evidence="not_experimental"
repeat_region 36551. .36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region 37585. .38254
```

```
repeat_region /note="U1PA15 repeat: matches 5480. .6157 of consensus"
38265. .38561
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 39214. .39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region 39489. .39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region 40050. .40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region 41057. .41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region 41296. .41460
/note="U2 repeat: matches 2569. .2730 of consensus"
repeat_region 41506. .41944
/note="U2 repeat: matches 1916. .2416 of consensus"
repeat_region 42388. .42698
/note="U2 repeat: matches 1448. .1779 of consensus"
repeat_region 44193. .44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region 44600. .44956
/note="U1R16A repeat: matches 90. .445 of consensus"
repeat_region 45240. .45300
/note="MER8A repeat: matches 37. .97 of consensus"
repeat_region 45798. .45909
/note="U1R41 repeat: matches 90. .192 of consensus"
repeat_region 46826. .46871
/note="23 copies 2 mer gt 97% conserved"
repeat_region 50684. .50969
/note="AluSq repeat: matches 1. .287 of consensus"
repeat_region 50980. .51291
/note="AluX repeat: matches 1. .312 of consensus"
repeat_region 52222. .52519
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region 54065. .54260
/note="U1M4 repeat: matches 3865. .4055 of consensus"
repeat_region 54261. .54432
/note="FAM repeat: matches 2. .167 of consensus"
repeat_region 54433. .54629
/note="U1M4 repeat: matches 3652. .3865 of consensus"
repeat_region 54648. .54862
/note="U1R41 repeat: matches 11. .217 of consensus"
repeat_region 54863. .55236
/note="MLT1A1 repeat: matches 1. .365 of consensus"
repeat_region 55237. .55700

Alignment Scores:
Pred. No.: 1.42e-43 Length: 165110
Score: 1099.00 Matches: 210
Percent Similarity: 98.60% Conservative: 1
Best Local Similarity: 98.13% Mismatches: 3
Query Match: 97.52% Indels: 0
DB: 9 Gaps: 0

US-09-595-947E-10 (1-214) x AL450311 (1-165110)
QY 1 MethPrpGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 30325 ATGAGCGCTCAACCTCGGGTGGCGCCACTGTCCAAAGAGACCCGTGAACGAGCGTCC 30266
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
DB 30265 TTCCCAAGAGCTCGAAGAGAGAGTGAACGACCCCAAGTCCGCCGCCAGCCCACT 30206
QY 41 ArgThrProGlyAsnGlyAlaGlnValGluGluGlyCysArgGlyAlaProArgGly 60
DB 30205 CGCACACGGGGGAACTGGCGAGAGCGGAGAGGAGAGGAGCTCCAGAGGAGGAGAG 30146
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerGlyGln 80
DB 30145 CTCGGGACGCGCGGGGAGACGACGCGGCTTAAGACGAGTGGCACTGAGCAAGCAG 30086
QY 81 ArgArgSerArgArgGlyLeuAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
```

Db	30085	CGACGGAGTGGCGGAAAGAGGCCAACGACCGGAGCGCAATGGATGGACCAACCTTCAC	30028
Qy	101	SerialaenaphalaleuAatrglyValleuProthrPheProaaphapalalytleuThr	120
Db	30025	TGGCAGCTGAGCAGCGCCCTCGCGGGGTCTCTGCGCACCTTCCACAGCAGCGGAAGCTCAC	29966
Qy	121	LysIleleuThrLeuAatgPheAlaHisAsnTyrIleThrPalaLeuThrGlnThrLeuAsy	140
Db	29965	AAGATTCAGAGACGTGCGCTTTCGCCACAACTACATCTGGGCGCTGACTCAAAAGCTTGGC	29906
Qy	141	IlealaaphasiserleuTyralaLeuGluProProalaProHisCyGlyGluLeuGly	160
Db	29905	ATACGCGACCACTGTACGCGCTGAGCGCGCGCGCACATCGCGGGAGCTGGGG	29846
Qy	161	SerProGlyGlyProProGlyAspTrpGlySerLeuTyserProValSerGlnaGly	180
Db	29845	AGCCAGACCGGCTTCCCGGGGACATGGGAGGTCCCTCACTACCCAGCTCCAGCGCTGGC	29786
Qy	181	SerleuSerProAlaAlaaserleuGluAlaProProGlyLeuLeuGlyAlaThrSer	200
Db	29785	AGCTTAGTCCCGCGCGGTCTGCTGAGAGAGAGACCGCGGCTGCTGGGGGCGACACTTTTCC	29722
Qy	201	AlaCysleuSerProGlySerleuAlaPheSerAspPheLeu	214
Db	29725	GCTTGCTTGAAGCCAGGACATGTCTTCTCAGATTTTCTG	29684
RESULT 4			
AC021954/C			
LOCUS	AC021954	173341 bp	DNA linear
DEFINITION	Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT		
SEQUENCE	24 unordered pieces.		
ACCESSION	AC021954		
VERSION	AC021954.3	GI:7417809	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 173341)		
TITLE	Biren,B., Lincon,L., Nusbaum,C. and Lander,E.		
JOURNAL	Homo sapiens chromosome 10, clone RP11-57E12		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 173341)		
Biren,B., Lincon,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Bara,N., Beckertly,R., Beda,F., Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,G., Castle,A., Choepey,Y., Colangelo,M., Collins,S., Colliore,A., Cooke,P., Dextrallano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J., Ferreira,P., Fitzhugh,W., Forreest,C., Gage,D., Galagan,L., Galdyna,S., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,D.C., Johnson,R., Jones,C., Kam,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marguis,N., McBum,P., McGuck,A., McKernan,K., McNetters,R., Meldrum,J., Menes,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,I.M., Peterson,K., Pletier,N., Pisan,C., Pollatz,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Titrill,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.			
TITLE	Direct Submission		
JOURNAL	Submitted (32-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA		
REFERENCE	3 (bases 1 to 173341)		
AUTHORS	Biren,B., Lincon,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Bara,N., Baetien,V., Beda,F., Boguslavsky,L., Bouhgalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A., Choepey,Y., Colangelo,M., Collins,S., Colliore,A., Cooke,P., Dextrallano,K., Dewar,K., Diaz,J.S., Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Galdyna,S., Ginde,S., Goyette,M., Graham,L., Grand-Pierre,N., Grant,G., Hagos,B., Heatford,A., Horton,L., Howland,D.C., Iliev,I., Johnson,R., Jones,C., Kam,L., Karatas,A.,		

TITLE
JOURNAL
COMMENT

```

Klein, J., LaRoque, K., Lamazeres, R., Landers, T., Lehoczy, J.,
Levine, R., Lien, C., Liu, G., Locke, K., Macdonald, P., Margis, N.,
McCarthy, M., McKwan, P., McGurk, A., McKernan, K., McPheters, R.,
Meldrum, J., Menues, L., Mhova, T., Miranda, C., Mlenga, V., Morrow, J.,
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,
Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
Strange-Thomann, N., Stojanovic, N., Sudrmanian, A., Talamas, J.,
Teisleye, S., Theodode, J., Tirelli, A., Travers, M., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,
Young, G., Zainoun, J., Zimmer, A. and Zody, W.

Submitted (24-NG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Apr 5, 2000 this sequence version replaced gi:6984451.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L5931
Center clone name: 57_E.12

----- Summary Statistics
Sequencing vector: MJ3; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 161190 bases at least Q40
Consensus quality: 166837 bases at least Q30
Consensus quality: 168995 bases at least Q20
Insert size: 176000; agarose-fp
Quality coverage: 3.7 in Q20 bases; agarose-fp
Quality coverage: 3.8 in Q20 bases; sum-of-contigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
1575: contig of 1574 bp in length
1575 1674: gap of 100 bp
1675 3043: contig of 1369 bp in length
3044 3143: gap of 100 bp
3144 5889: contig of 2746 bp in length
5890 5980: gap of 100 bp
5980 8979: contig of 2990 bp in length
8980 9079: gap of 100 bp
9080 13674: contig of 4595 bp in length
13675 13775: gap of 100 bp
13775 18831: contig of 5057 bp in length
18832 18931: gap of 100 bp
18932 23526: contig of 4595 bp in length
23527 23626: gap of 100 bp
23627 27386: contig of 3760 bp in length
27387 27486: gap of 100 bp
27487 32572: contig of 5086 bp in length
32573 32672: gap of 100 bp
32673 38632: contig of 5960 bp in length
38633 38732: gap of 100 bp
38733 43735: contig of 5003 bp in length
43736 43835: gap of 100 bp
43836 49020: contig of 5185 bp in length
49021 49120: gap of 100 bp
49121 53660: contig of 4540 bp in length
53661 53760: gap of 100 bp

```

*	53761	59544:	contig of 5784 bp	in length
*	59645	59644:	gap of 100 bp	
*	59645	66395:	contig of 6751 bp	in length
*	66396	66395:	gap of 100 bp	
*	66496	74696:	contig of 8201 bp	in length
*	74697	74796:	gap of 100 bp	
*	74797	81326:	contig of 6530 bp	in length
*	81327	81426:	gap of 100 bp	
*	81427	89446:	contig of 8020 bp	in length
*	89447	89546:	gap of 100 bp	
*	89547	96559:	contig of 6113 bp	in length
*	96460	96559:	gap of 100 bp	
*	96560	106498:	contig of 9939 bp	in length
*	106499	106598:	gap of 100 bp	
*	106599	116986:	contig of 10388 bp	in length
*	116987	117086:	gap of 100 bp	
*	117087	128890:	contig of 11804 bp	in length
*	128891	128990:	gap of 100 bp	
*	128991	147290:	contig of 13300 bp	in length
*	147291	147390:	gap of 100 bp	
*	147391	173441:	contig of 22551 bp	in length

```

source
1. 173341
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="10"
   /map="10"
   /clone_id="PPI1-57E12"
   /clone_1fb="RPCI-11 Human Male BAC"
misc_feature
1. 1574
   /note="assembly_fragment"
misc_feature
1675..3043
   /note="assembly_fragment"
misc_feature
3144..5889
   /note="assembly_fragment"
misc_feature
5990..8879
   /note="assembly_fragment"
misc_feature
9080..13674
   /note="assembly_fragment"
misc_feature
13775..18831
   /note="assembly_fragment"
misc_feature
18932..23526
   /note="assembly_fragment"
misc_feature
23627..27386
   /note="assembly_fragment"
misc_feature
/note="assembly_fragment
clone_end:SP6
vector_side:left"
27487..32572
   /note="assembly_fragment"
misc_feature
32673..38632
   /note="assembly_fragment"
misc_feature
38733..43735
   /note="assembly_fragment"
misc_feature
43836..49020
   /note="assembly_fragment"
misc_feature
/note="assembly_fragment"
49121..53660
   /note="assembly_fragment"
misc_feature
53761..59544
   /note="assembly_fragment"
misc_feature
59645..66395
   /note="assembly_fragment"
misc_feature
66496..74636
   /note="assembly_fragment"
misc_feature
74797..81326
   /note="assembly_fragment"
misc_feature
81427..89446
   /note="assembly_fragment"
misc_feature
89547..96459
   /note="assembly_fragment"
misc_feature
96560..106498
   /note="assembly_fragment"
misc_feature
106599..116986

```

```
misc_feature /note="assembly_fragment
117087.128890
/note="assembly_fragment
clone_end:t7
vector_side:right"
128991.147290
/note="assembly_fragment
147391.173341
/note="assembly_fragment
```

Alignment Scores:

Pred. No.:	4.54e-43	Length:	173341
Score:	1089.00	Matches:	209
Percent Similarity:	98.13%	Conservative:	1
Best Local Similarity:	97.66%	Mismatches:	4
Query Match:	96.63%	Indels:	0
DB:	2	Gaps:	0

US-09-595-947E-10 (1-214) X AC021954 (1-173341)

1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20

Db 6765 ATGACGCTCAACCTCGGTGCGCCCACTGTCCAAGTGACCCGTGAGACGGAGCGTCC 6706

21 PheProArgAlaSerGluaspGluValThrCysProThrSerAlaProProSerProThr 40

Db 6705 TTCCCAGAGCCTCGGAGACGAAGTGACCTGCCCCACGTCGGCCCCCGCCAGCCCCACT 6646

QY 41 ArgThrProGlyAsnCySAIAGIAGIUGIUGIYCYSARGGIYAIAProARGLys 60

Db 6645 CGCACACGGGGAACTGCCGAGAGGCGGAGAGGGAGGCTGCCGAGGGGCCCGAGGAAG 6586

61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80

Db 6585 CTCGGGCA CGGCGCGGGGACGCAGCCGGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 6526

81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

Db 6525 CGACGGAGTCCGGCGAAGAAGGCCAACGACCGCGAGCGCAATCGAATGCACAACTCAAC 6466

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

Db . 6465 TCGGCACTGACGCCCCCTGCGCGGTGTCTCTGCCACCTTCCAGACGACGGGAAGTACCC 6406

121 LysIleGluThrLeuArgpheAlaHisAsnTyrIleTrpAlaLeuThrGlnThrLeuArg 140

Db 6405 AAGATCGAGACGCTGCGCTTCGCCCAACAATACTACATCTGGCGCGTGACTCAACGCTGCC 6346

141 ILEIAspRHisSerLeuTYRAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160

D5 6345 ATAGCGACCA CAGCTGTACGCGCTGGAGCCGCCGCGCACTGCGGAGCTGGC 6286

161 SerProGlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAlaGly 180

Db 6285 AGCCCAAGCGTTCCTCCCGGACATGGGGTCCCTACATCCCAAGTCTCCAGCTGGC 6226

181 SerLeuSerProAlaIaSerLeuGluGluArgProGlyLeuLeuGlyAlaInrSerSer 200

D5 6225 AGCCTGAGTCCCGCCGCGTCGCTGGAGGAGCGACCCCGGCTGCTGGGGCCACCTTATCC 6160

201 AlacysleuserProGlyserLeuAlapheserAspPheneu 214

Db 6165 GCCCTGCTTGAAGCCCAAGCAGTCTGGCTTCTCAGATTCTG 6124

RESULT 5
A91167

LOCUS	A91167	1491 bp	DNA	linear	PA1 22-JAN-2000
DEFINITION	Sequence 1 from Patent WO9827206.				

ACCESSION	A91167
VERSION	A91167.1
GT:6740202	

KEYWORDS : Pattern and SOURCE

ORGANISM *Rattus* sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1491)

AUTHORS Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.

JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;
ICARD LIEPKALNS CHRISTINE (FR); MALLET JACQUES (FR)

FEATURES
Location/Qualifiers

CDS

1..1491
/organism="Rattus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10118"
459..1103
/note="unnamed protein product"
/codon_start=1
/protein_id="GI:6740203"
/translation="MAFPLDAPTVQVQSEIQPPFASDHEVLSNSTPPSPITVPR
DCESEADDCRGTSLKLRARGRNRPSELSQKRSRRKANDRRNRNMNLSA
LDALGVLPPTPDDAKLTETLRFAHYIWMLTQTLRIADISFYGPPEPVCCEIGS
FGGSSGWMGSIVSQAGSLSPASLEEFGLQVPSPSCLPLGLVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN

Alignment Scores:

Pred. No.: 3.02e-33 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 6 Gaps: 2

US-09-595-947E-10 (1-214) x A91167 (1-1491)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCTCATCTCTTGATGCGCCCAACATCCAAAGTGTCCCAAGAGCCAGCAACCC 518
QY 21 PheProGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCCCGAGGCTTCGGACCAAGAGTGTCTCAATTCACCCCACTTACCCCACT 578
QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGACCGAGGAGCTGCTCCGACAGAGAGGAGTCTGCGAGGAGCAATCGAGAG 638
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCCTGTGGCGCGCGGAGGCGCAACAGCCCAAGAGCGAGTTCAGTGAAGAG 698
QY 81 ArgArgSerArgArgGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 699 CGACGAGCGCGCGAGAGAGGCAACAGCCGAGGAGCGCAACGATGCACAACTTAA 758
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 759 TCCGCGCTGAGAGCGCTGCGGTCTCTCCGCACTTCCCGAGAGAGCCCAACTTACA 818
QY 121 LysIleGluThrLeuArgPheAlaHisValTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 819 AAGATCGAGACCTCTCGCTTCCGCCCAACTATTTGGGCACTGACTCAGACGCTGGGC 878
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 879 ATAGGAGACCAAGCTTCTACGAGCCCGAGCCCTCTGCCC--TGTGGAGAGCTGGGA 935
QY 161 SerPro--GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
DB 936 AGCCCGGAGGAGGCGCTCCAGCGGAGCACTGAGGCTCTATCTACTCCCACTTCCCAAGCT 995
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 996 GTTACCTGAGGCCCAAGCCCTATTGGAGGAGTTCCTGGCTCAGGTGCCAGCTCC 1055

QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 1056 CCATCTGTGTGCTCCCGGAGCACCCTGTGTTCTTCAAGCTTCTTG 1100

RESULT 6
BD023626 1491 bp DNA linear PAT 27-AUG-2002
LOCUS Polypeptide belonging to the family of basic helix-loop-helix
DEFINITION (bHLH) family and nucleic acid sequence corresponding thereto.

ACCESSION BD023626.1 GI:22564849

VERSION JP 2001510464-A/1.

KEYWORDS Rattus sp.

SOURCE Rattus sp.

ORGANISM Rattus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1491)

AUTHORS Liepkalns, C.I., Mallet, J. and Ravassard, P.

JOURNAL Polypeptide belonging to the family of basic helix-loop-helix
(bHLH) family and nucleic acid sequence corresponding thereto
Patent: JP 2001510464-A 1 31-JUL-2001;

COMMENT

OS Rattus sp. (rat)
PN JP 2001510464-A/1
PD 31-JUL-2001
PF 19-DEC-1997 JP 1998527415
PR 19-DEC-1996 FR 96/15651
PI CHRISTINE ICARD LIEPKALNS, JACQUES MALLET, PHILIPPE RAVASSARD PC
C12N15/09,
C12N15/00,A61K37/02
CC Strandedness: Single;
CC Topology: Linear;
FH Key

FEATURES

source

1..1491
/organism="Rattus sp."
/mol_type="genomic DNA"
/db_xref="taxon:10118"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN

Alignment Scores:

Pred. No.: 3.02e-33 Length: 1491
Score: 849.00 Matches: 166
Percent Similarity: 83.72% Conservative: 14
Best Local Similarity: 77.21% Mismatches: 33
Query Match: 75.33% Indels: 2
DB: 6 Gaps: 2

US-09-595-947E-10 (1-214) x BD023626 (1-1491)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGGCGCTCATCTCTTGATGCGCCCAACATCCAAAGTGTCCCAAGAGAGCCAGCAACCC 518
QY 21 PheProGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTTCCCGAGGCTTCGGACCAAGAGTGTCTCAATTCACCCCACTTACCCCACT 578
QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGACCGAGGAGCTGCTCCGACAGAGAGGAGTCTGCGAGGAGCAATCGAGAG 638
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCCTGTGGCGCGCGGAGGCGCAACAGCCCAAGAGCGAGTTCAGTGAAGAG 698
QY 81 ArgArgSerArgArgGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 699 CGACGAGCGCGCGAGAGAGGCAACAGCCGAGGAGCGCAACGATGCACAACTTAA 758

QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLeuThr 120
 Db 759 TCCGGCTGATGAGCTGGCGGTGCTGCGCCACTTCCGAGTACGCAACTTACA 818
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
 Db 819 AAGATCGAGACCTCGCTTCCGCCACATTCATTGGGCACTGACAGCGCTGCC 878
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaPheProHisCysGlyValLeuGly 160
 Db 879 ATGGGAGACCAAGCTTCTACGCGCCCGACCCCTGTGCC---TGTGGGAGCTGGGA 935
 QY 161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
 Db 936 AGCCGGGAGGGGGCTCCAGCGGCGACTGGGCTCATCTACTCCCAAGTTCCCAAGCT 995
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 996 GGTACCTGAGCCCAACAGCTCATTTGAGAGAGTTCCCTGGCTCGAGGTCGCCAGCTCC 1055
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 1056 CCATCTGTCTCTCTCCCGGACCTGTGTTCTCAGACTTCTTG 1100
 RESULT 7
 RNRRLAXT 1491 bp mRNA linear ROD 06-MAY-1997
 LOCUS R. norvegicus mRNA for transcriptional regulator, Relax.
 DEFINITION Y10619.1 GI:2072737
 VERSION Y10619.1 GI:2072737
 KEYWORDS Relax; transcriptional regulator.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1
 AUTHORS Ravassard, P., Chatail, F., Mallet, D. and Icard-Jilekals, C.
 TITLE Relax, a novel rat bHLH transcriptional regulator transiently
 expressed in the ventricular proliferating zone of the developing
 central nervous system
 JOURNAL J. Neurosci. Res. 48 (2), 146-158 (1997)
 MEDLINE 97276390
 PUBMED 9130143
 REFERENCE 2 (bases 1 to 1491)
 AUTHORS Ravassard, P.
 TITLE Direct Submission
 JOURNAL Submitted (20-JAN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVI,
 Hopital de la Pitie Salpêtrière, 83 Bd. de l'Hopital, F-75013
 Paris, FRANCE
 FEATURES
 source location/Qualifiers
 1..1491
 /organism="Rattus norvegicus"
 /mol_type="mRNA"
 /strain="Wistar"
 /db_xref="taxon:10116"
 /clone_1fb="RT-PCR"
 459..1103
 /function="transcriptional regulator"
 /codon_start=1
 /product="Relax"
 /protein_id="CA71630.1"
 /db_xref="GI:2072738"
 /db_xref="SPTREML:008718"
 /translation="MAPHPUDAPTIVQSETOQPPGASDHEVLSNTPSPETLVR
 DSEAGADCKGRTSLRKARGGRRPKSELISKRRRRKKANDRRNNHNSA
 IDALGVLPDPDAKLTLETRFANHYIMALTQRLADHSFGEPEPVCGEIGS
 PGGSGSDGSIYSPVQAGSLSPFASLEFPGIQVPSPPCILPGLIVFSDFL"
 BASE COUNT 307 a 487 c 413 g 284 t
 ORIGIN
 Alignment Scores:

Pred. No.: 3,02e-33 Length: 1491
 Score: 849.00 Matches: 166
 Percent Similarity: 83.72% Conservative: 14
 Best Local Similarity: 77.21% Mismatch: 33
 Query Match: 75.33% Indels: 2
 DB: 10 Gaps: 2
 US-09-595-947E-10 (1-214) x RNRRLAXT (1-1491)
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 459 ATGGGCTCATCTCTTGATGAGCGCCACATTCAGAGTGTCCAGAGACCCAGACACC 518
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 Db 519 TTTCCGAGAGCTCGACAGACAGAGTGTCTCAGTTCCATTCCACCCCTGAGCCCACT 578
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyCysArgGlyAlaProArgLys 60
 Db 579 CTCGTACCGAGGAGCTGCTCCGAGACAGAGGTAAGTCCGAGGAGCATCGAGAG 638
 QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 639 CTCGTGCGCGCGCGGAGGCGGCAAGCCCAAGAGCGAGTTGCACTGAGCAAGCAG 698
 QY 699 CGACGAGCCGCGCGAGAGAGCCACAGACCGGCGCAACCCGATCAGACACTTAA 758
 QY 81 ArgArgSerArgArgGlyValAlaAspArgGluArgAsnArgMetHisAspLeuAsn 100
 Db 699 CGACGAGCCGCGCGAGAGAGCCACAGACCGGCGCAACCCGATCAGACACTTAA 758
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLeuThr 120
 Db 759 TCCGGCTGATGAGCTGGCGGTGCTGCGCCACTTCCGAGTACGCAACTTACA 818
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
 Db 819 AAGATCGAGACCTCGCTTCCGCCACATTCATTGGGCACTGACAGCGCTGCC 878
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaPheProHisCysGlyValLeuGly 160
 Db 879 ATGGGAGACCAAGCTTCTACGCGCCCGACCCCTGTGCC---TGTGGGAGCTGGGA 935
 QY 161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
 Db 936 AGCCGGGAGGGGGCTCCAGCGGCGACTGGGCTCATCTACTCCCAAGTTCCCAAGCT 995
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 996 GGTACCTGAGCCCAACAGCTCATTTGAGAGAGTTCCCTGGCTCGAGGTCGCCAGCTCC 1055
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 1056 CCATCTGTCTCTCTCCCGGACCTGTGTTCTCAGACTTCTTG 1100
 RESULT 8
 AC127817 258815 bp DNA linear HTG 19-NOV-2002
 LOCUS Rattus norvegicus clone CH230-259G16, WORKING DRAFT SEQUENCE, 3
 DEFINITION Rattus norvegicus
 UNORDERED PIECES.
 AC127817
 AC127817.3 GI:25077905
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 REFERENCE 1 (bases 1 to 258815)
 AUTHORS Muzny, D., Marie, J., Metker, M., Lee, J., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Albrooks, S., Amin, A., Angilano, D.,
 Anyalebech, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behan, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Bunay, C., Burch, P., Burrell, K., Calderon, E.,

Cardenas V., Carter K., Cavazos, I., Censar H., Center A.,
Chacko J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
Devigila, M.L., Davis, S., Davy-Carroll, L., De Aranda, C., Decker, D.,
Delgado, O., Denison, S., Dermoco, C., Ding, Y., Dinh, H., Ditych, K.,
Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
Fraser, C.M., Gabisi, A., Galla, R., Garcia, A., Garner, T., Garza, M.,
Gellergoizis, E., Geer, K., Gint, R., Grady, M., Guerin, M., Guevara, W.,
Gunaarane, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Hernandez, R., Hines, S., Hladin, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huliy, S., Hune, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A.,
Karpaty, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
Kovis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
Liu, J., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshewa, L., Louisseg, H., Lozdo, R.J., Lu, X., Ma, T.,
Maneshwari, M., Mahindaratne, M., Mamoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
Mahniny, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Mirza, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Mundaas, M., Murphy, M., Natr, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwoketlemah, O., Okwomu, G., Olarnunagsoon, A., Pal, S., Parks, K.,
Paeternack, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
Plommer, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L.,
Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
Riley, B., Rellay, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
Rivers, C., Rodkey, T., Rojao, A., Rose, M., Rose, R., Riz, S.T.,
Sanders, M., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D.,
Sned, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Sytek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejo, Z., Umani, K.,
Valas, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J.,
Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Willson, R., Wleczek, R., Wooden, H., Wolley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhao, S., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G., and Gibbs, R.A.

Direct Submission
2 (bases 1 to 258815)
Unpublished
Worley K.C.

Direct Submission
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258815)
Rat Genome Sequencing Consortium.

Direct Submission
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 19, 2002 this sequence version replaced gi:23912578.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.tgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM

```

Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: GZXS
Center clone name: CH230-259G16
----- Summary Statistics -----
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228243; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contig. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*      1   255979: contig of 255979 bp in length
*      *   255980   256079: gap of unknown length
*      *   256080   257349: contig of 1270 bp in length
*      *   257350   257449: gap of unknown length
*      *   257450   258815: contig of 1366 bp in length.
*
*      Location/Qualifiers
*          1..258815
*              /organism="Rattus norvegicus"
*              /mol_type="genomic DNA"
*              /db_xref="taxon:10116"
*              /clone="CH230-259G16"
*
*      misc_feature
*          1..1609
*              /note="wgs end-extension
*                  clone_end:Sp6"
*
*      misc_feature
*              complement(2865..3659)
*              /note="clone boundary
*                  clone_end:Sp6
*                  site:"
*
*      misc_feature
*              end_sequence:BZ169378"
*              224098..225157
*              /note="clone boundary
*                  clone_end:T7
*                  site:"
*
*      BASE COUNT      59351 a 56888 c 58131 g 56821 t 27624 others
*      ORIGIN
*          Alignment Scores:
*      Pred. No.:       2.94e-31           Length:        258815
*      Score:           849.00             Matches:         166
*      Percent Similarity: 83.72%          Conservative:    14
*      Best Local Similarity: 77.21%       Mismatches:     33
*      Query Match:      75.33%            Indels:        2
*      DB:                2               Gaps:           2
*
US-09-595-947E-10 (1-214) x AC127817 (1-258815)
OY      1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db      247860 ATGGCGCTCATCCCTTGATGCGGCCACCAATCCAAAGTGTCCCAAGAAGACCAACC 247919
OY      21 PheProAlaGalaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db      247920 TTTCCTGGAGCTCGGACCAAGAAATGTCTCAATTCCAACCCCACTTAGCCCCACT 247979
OY      41 ArgThrProGlyAsnCyAlaGluIuaIagluGluGlyGlyCyAargGlyValaProArGlys 60
Db      247980 CTCGACCGAAGGATGCTCCGAACAAGACAGATGACTGCGAAGGAGCATCGAAGAG 248039
OY      61 LeuAlaGlaArgArgGlyGlyAArgSerAArgProLysSerGluLeuAlaLeuSerLysGln 80

```

```

Db      248040 CTCCTGCGCGGCGGAGGCGGCAACAGCCCAAGAGAGTGGCACTGAGCAAGCG 248099
Qy      81  ATGAGSerserArgArgLyLysAlaAsnAspArgGluArgAsnArgMethIsaAspleuAen 100
Db      248100 CGACGAGCGCGCGCAAGAGGCGCAACGACCGGAGCGCAACCGCATGCAACACTTAA 248159
Qy      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db      248160 TCCGCGCTGATGCGCTGCGCGGCTGCTGCGCACTTCCCGATGACCGCAAACTTACA 248219
Qy      121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIeThrAlaLeuThrGlnThrLeuArg 140
Db      248220 AAGATCGAAGCCCTGCGCTTCCGCCCAACATCTATTGGGCACTGACGAGCTGCGC 248279
Qy      141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisGlyGluLeuGly 160
Db      248280 ATAGCGGACCAACAGCTTCTACGCGCCCGGAGCCCTCTGTGCCC---TGTTGGAGCTGGGA 248336
Qy      161 SerPro---GlyGlyProProGluLysPrtGlySerLeuTyrlSerProValSerGlnAla 179
Db      248337 ACCCGGAGGAGGAGGCTCCAGCGCGGAGCTGGGCTCTACTACTCCCGAGTTCCCAAGCT 248396
Qy      180 GlySerLeuSerProAlaLysLeuGluGluArgProGlyLeuLeuGluValaThrSer 199
Db      248397 GGTAGCCTGAGGCCCAACAGCTCTATTGAGGAGTTCCCTGCGCTGCAAGGTGCCCAAGCTCC 248456
Qy      200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      248457 CCATCTGTCTCTCTCCCGGAGCCCTGTGTCTCTCAGACTTCTTG 248501

```

```

RESULT 9
LOCUS      AX698801      861 bp      DNA      linear      PART 02-APR-2003
DEFINITION Sequence 7 from Patent WO02086107.
ACCESSION  AX698801
VERSION     AX698801.1 GI:29499589
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  Wobus,A.M., St-Onge,L., Blyszczuk,P. and Hoffmann,U.
            A method for differentiating stem cells into insulin-producing
            cells
JOURNAL     Patent: WO 02086107-A 7 31-OCT-2002;
            Deutscher Forschungsbund fuer Entwicklungsbio-logische Forschung
            (DFB); INSTITUT FUR PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
            (DFP)
FEATURES
            source
            1..861
            /location="Qualifiers"
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            160..804
            /note="unnamed protein product"
            /codon_start=1
            /protein_id="CAD87442.1"
            /db_xref="GI:29499590"
            /translation="MAHPPLDALVITQVSPETQPPFGASDHEVLSSNTPSPPTLIPR
            DCEAGVDCRGSRKLARAGRGRRNPKSELISQRRSRKRANDRRNRNMINNSA
            IDALRGVLPDPDAKLRTIETLRFAHNTLTSLRADSFGYGPPEPVCGEIGS
            PGGSNGMDGSIYSPSOGNLSPTASLEPFLQVPSPSYLLPGLVPSDFL"
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN

```

```

Alignment Scores:
Pred. No.:      5,32e-32      Length:      861
Score:          819.00      Matches:      163
Percent Similarity: 82.33%      Conservative: 14
Best Local Similarity: 75.81%      Mismatches: 36
Query Match:    72.67%      Indels:      2

```

```

DB:      6      Gaps:      2
US-09-595-947E-10 (1-214) x AX698801 (1-861)
Qy      1  MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db      160 ATGGCGCTCATCTCTTGGATGCGCTCACCATCAGATGTCTCCAGAGACACACAACT 219
Qy      21  PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db      220 TTTCGCGAGCTCGGACACAGAACTGCTCAGTTCCATTCCACCCTGAGCCCACT 279
Qy      41  ArgThrProGlyAsnGlyAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
Db      280 CTGATACCTTAGGAGCTGCTCCGAGCAAGAGTGAGTGTCTGCGAGGAGCTCGAGGAAG 339
Qy      61  LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db      340 CTCGCGCGCCGAGCGGAGGCGGCAACAGGCCCAAGAGCGAGTTGGCACTGCAACAG 399
Qy      81  ATGAGSerserArgArgLyLysAlaAsnAspArgGluArgAsnArgMethIsaAspleuAen 100
Db      400 CGAAGAGCGCGCGCAAGAGGCGCAATGTCGGAGCGGCAATCGCATGCAACACTTAA 459
Qy      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120
Db      460 TCGCGCGTGAAGTGGCTGCGCGGCTGCTGCTGCGCACTTCCCGATGAGCGCCAACTTACA 519
Qy      121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIeThrAlaLeuThrGlnThrLeuArg 140
Db      520 AAGATCGAAGCCCTGCGCTTCCGCCCAACATCTATTGGGCACTGACGAGCTGCGC 579
Qy      141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisGlyGluLeuGly 160
Db      580 ATAGCGGACCAACAGCTTCTACGCGCGGAGCCCTGTGCCC---TGTTGAGAGCTGGGG 636
Qy      161 SerPro---GlyGlyProProGluLysPrtGlySerLeuTyrlSerProValSerGlnAla 179
Db      637 AGCCCGGAGGAGGCTCCAGCGGAGCTGGGCTCTACTACTCCCGAGTCCCAAGCG 696
Qy      180 GlySerLeuSerProAlaLysLeuGluGluArgProGlyLeuLeuGluValaThrSer 199
Db      697 GGTAGCCTGAGGCCCAACAGCTCTATTGAGGAGTTCCTGCGCTGCAAGGTGCCCAAGCTCC 756
Qy      200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      757 CCATCTATCTCTCTCCCGGAGCACTGTGTCTCAGACTTCTTG 801

```

```

RESULT 10
LOCUS      MMU76208      861 bp      DNA      linear      ROD 05-FEB-1997
DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.
ACCESSION  U76208
VERSION     U76208.1 GI:1815654
KEYWORDS
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  Sommer,L., Ma,Q. and Anderson,D.J.
            neurogenins, a novel family of atonal-related bHLH transcription
            factors, are putative mammalian neuronal determination genes that
            reveal progenitor cell heterogeneity in the developing CNS and PNS
JOURNAL     Mol. Cell. Neurosci. 8 (4), 221-241 (1996)
MEDLINE     97153565
PUBMED      9000438
REFERENCE   2  (bases 1 to 861)
AUTHORS     Sommer,L., Ma,Q. and Anderson,D.J.
TITLE       Direct Submissio
JOURNAL     Submitted (24-OCT-1996) Biology 216-76, California Institute of
            Technology, Howard Hughes Medical Institute, Pasadena, CA 91125,
            USA

```

COMMENT On Feb 5, 1997 this sequence version replaced gi:166911.
FEATURES Location/Qualifiers

source

1..861

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

160..804

/gene="ngn3"

160..804

/gene="ngn3"

/function="bHLH transcription factor"

/note="NGN3"

/codon_start=1

/product="neurogenin 3"

/protein_id="AA053029.1"

/db_xref="GI:166912"

/translation="MAHPPLDALITQVSPETQPPPGASDHEVLSNSNTPSPPTIPR
DSEAEVDCRGTSRKLARRGRNRPKSELALSKQRARRKADRRNRNMINSA
LDALRGVLPFPDDAKLTETLRPAHNYIMALTOTLIADHSFYGPPEPVPCGELGS
PGGSGNDWGSIVSPVQAGNISPTASLEBPQLQVPSPSYLLPGALVFDPL"

BASE COUNT

182 a 274 c 250 g 155 t

ORIGIN

Alignment Scores:

Pred. No.: 5,32e-32 Length: 861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2

US-09-595-947e-10 (1-214) x MMU76208 (1-861)

QY 1 MetThrProGlnProSeGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 160 ATGGCGCCTCATCTCCCTTGATGGCTGCACATCAAGTGTCCCGAGAGACACAAACCT 219
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 220 TTTCCCGAGCCTCGAGACCAAGAGTCTCAGTTCATTCACATCCACCTTACGCCCCACT 279
QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgLys 60
DB 280 CTCATACCTAGAGACTGCTCCGAGCAGAAAGTGGGTGACTGCCGAGGACCTCGAGGAAG 339
QY 61 LeuAlaGluAlaArgArgGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 340 CTCGCGCCCGACGCGGAGGCGGCAACAGCCCAAGAGGAGTTCGACCTCAGCAAAACAG 399
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 400 CCAAGAGCCCGGCAAGAGGCGCAATGATCGGAGCGCAATGCGACAACTCCAAC 459
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 460 TCGGGCGCTGAGCGCGTGGCGGTGCTGCCCACTTCGCGGATGAGCCAAACTTAA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 520 AAGATCGAAGCCTGCTGCGCCCAACTATCTGCGCACTGACTCAGAGCTGCGC 579
QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisValGlyLeuGluArg 160
DB 580 ATAGCGACCAAGCTTCTATGCGCCGAGGCGCCCTGTGCC--GTGGAAGAGCTGGGG 636
QY 161 SerPro--GlyGlyProProGlyAspTyrGlySerLeuThrSerProValSerGlnAla 179
DB 637 AAGCCCGAGAGTGGCTCCAAAGGAGCTGGGCTCATATCATCTCCCACTTCCAAAGG 696
QY 180 GlySerLeuSerProAlaAsnSerLeuGluGluArgProGlyLeuGluAlaThrSer 199
DB 697 GGTAACCTAGGCCCCAGCGCTCATTTGAGGAATTCCTGCGCTGAGGCTGCCAGCTCC 756

QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 757 CCATCTATCTGCTCCCGGAGCACTGGTCTTCTAGACTTCTTG 801

RESULT 11.

AF364300

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

ORGANISM

SOURCE

REFERENCE

AUTHORS

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

JOURNAL

AUTHORS

REFERENCE

TITLE

MRNA

CDS

1..1861

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

<1093..>1737

/product="neurogenin 3"

1093..1737

/codon_start=1

/product="neurogenin 3"

/protein_id="AAK5058.1"

/db_xref="GI:13937129"

/translation="MAHPPLDALITQVSPETQPPPGASDHEVLSNSNTPSPPTIPR
DSEAEVDCRGTSRKLARRGRNRPKSELALSKQRARRKADRRNRNMINSA
LDALRGVLPFPDDAKLTETLRPAHNYIMALTOTLIADHSFYGPPEPVPCGELGS
PGGSGNDWGSIVSPVQAGNISPTASLEBPQLQVPSPSYLLPGALVFDPL"

BASE COUNT

397 a 560 c 537 g 367 t

ORIGIN

Alignment Scores:

Pred. No.: 1.05e-31 Length: 1861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 10 Gaps: 2

US-09-595-947e-10 (1-214) x AF364300 (1-1861)

QY 1 MetThrProGlnProSeGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 1093 ATGGCGCCTCATCTCCCTTGATGGCTGCACATCAAGTGTCCCGAGAGACACAAACCT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 1153 TTTCCCGAGCCTCGAGACCAAGAGTCTCAGTTCATTCACATCCACCCACTAGCCCACT 1212
QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgLys 60
DB 1213 CTCATACCTAGAGACTGCTCCGAGCAGAAAGTGGGTGACTGCCGAGGAGCTCGAGGAAG 1272
QY 61 LeuAlaGluAlaArgArgGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 1273 CTCGCGCCCGACGCGGAGGCGGCAACAGCCCAAGAGGAGTTCGACTCAGCAAAACAG 1332
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

Db 1333 CGAAGAGCGCGCGAGAGGCCAATGATCGGAGCGGCAATCGCATGCAACCTCAAC 1392
 QY 101 SerAlaLeuAspAlaLeuArglyValLeuProThrPheProAspAlaIleuLeuThr 120
 Db 1333 TCGGGCTGGATGCGCTGGCGGTGTCTCTCCCACTTCCCGATGACGCCAATCTTACA 1452
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPAlaLeuThrGlnThrLeuArg 140
 Db 1453 AAGATCGAGACCTGGCTTCCGCCCACTACATCTGGGCACTGACTCAGACGCTGCC 1512
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 Db 1513 ATAGGGACACACAGCTTCTATGCGCGGACCCCTGTCCTCC---TGTGAGAGCTGGGG 1569
 QY 161 SerPro---GlyGlyProProGlyAspTPGlySerLeuTyrSerProValSerGlnAla 179
 Db 1570 ACCCCCGAGGTGGCTCCACAGGGAGCTGGGCTCTATGTTACTTCCCACTTCCCAAGG 1629
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
 Db 1630 GGTAACCTGAGCCCGACGCGCTCATTTGAGGAATTCCTGGCTGAGGTCGCCAGCTCC 1689
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 Db 1690 CCATCTATCTCTCCCGGAGACACTGTGTCTCAGACTTCTTG 1734
 RESULT 12
 MMAT4B
 LOCUS MMAT4B 5567 bp DNA linear ROD 25-OCT-2000
 DEFINITION M.musculus MAT4B gene.
 ACCESSION Y09167
 VERSION Y09167.2 GI:11065737
 KEYWORDS bHLH; mammalian atonal homolog 4B; MAT4B; neurogenin 3.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1
 AUTHORS Cau, E., Gradwohl, G., Fode, C. and Guillemot, F.
 TITLE Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors
 JOURNAL Development 124 (8), 1611-1621 (1997)
 MEDLINE 97261963
 PUBMED 9108377
 REFERENCE 2
 AUTHORS Jacquemin, P., Duriaux, S.M., Jensen, J., Godfrand, C., Gradwohl, G., Guillemot, F., Maden, O.D., Carmeliet, P., Dwerchin, M., Collen, D., Rousseau, G.G. and Lemaigre, F.P.
 TITLE Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3
 JOURNAL Mol. Cell. Biol. 20 (12), 4445-4454 (2000)
 MEDLINE 20285449
 PUBMED 10825208
 REFERENCE 3
 AUTHORS Gradwohl, G.J.
 TITLE Direct Submission
 JOURNAL Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC, CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg, F-67404 ILKIRCH cedex, FRANCE
 REMARK 4 (bases 1 to 5567)
 REFERENCE Lemaigre, F.P.
 TITLE Direct Submission
 JOURNAL Submitted (01-AUG-2000) Lemaigre F.P., Hormone and Metabolic Research Unit, Louvain University Medical School, Avenue Hippocrate 75, box 7529, Brussels 1200, BELGIUM
 COMMENT On Oct 31, 2000 this sequence version replaced gi:1666087.
 FEATURES
 source
 1. 5567
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /strain="SV129"

/db_xref="taxon:10090"
 /clone="lambda clone 19c"
 /cell_line="D3"
 /clone_lib="embryonic stem cells"
 /clone_id="Genomic library No.62, Mbol partial digests cloned in BamHI"
 /dev_stage="embryos"
 misc_binding
 /bound_moiety="HNF6"
 3692
 misc_binding
 /bound_moiety="HNF6"
 4145..5567
 gene
 /gene="MAT4B"
 4145..4151
 TATA_signal
 /gene="MAT4B"
 /note="putative"
 4923..5567
 CDS
 /gene="MAT4B"
 /product="neurogenin 3"
 /codon_start=1
 /protein_id="CAA70366.1"
 /db_xref="GI:1666087"
 /db_xref="SWISS-PROT:P70661"
 /translation="MAHPIDALTTQVSPETQDPRPGASDHEVLVSNSNTPSPITLPR
 DSEAEVDCRGTSKLFAKRGRRRPSRLASFKSRKKANDRKNRNLNSA
 LDALRGVLPETPDAAKLTKIETLRPAHYVALTQTLADHSFYGPPEPVCGLS
 FGSGNCGMGSIVSPVSOQNLSPASLEEFPGIQVPSSPYLLFGALVPSDFL"
 BASE COUNT 1271 a 1549 c 1564 g 1183 t.
 ORIGIN
 Alignment Scores:
 Pred. No.: 2,79e-31 Length: 5567
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 10 Gaps: 2
 US-09-595-947E-10 (1-214) x MMAT4B (1-5567)
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 4923 ATGGCGCTCATCTCTGGATGCGCTCACCATCCAGTGTCCCGAGACACAACTCACT 4982
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 Db 4983 TTTCCCGAGAGCTTCGGACACACAACTGCTCAATTCACACCCCACTGACCCCACT 5042
 QY 41 ArgThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgLys 60
 Db 5043 CTCATACCTAGGAGCTGCTCCGACACAGAGGTGCTGCTCCGAGGACCTCGAGGAG 5102
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 5103 CTCGCGCCCGACGCGAGGGGCGCAACAGCCCAAGAGAGTGGCACTCAGCAAAACG 5162
 QY 81 ArgArgSerArgArgLysLeuAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 Db 5163 CGAAGAGCCGCGCAAGAGCCCAATGATGGAGCGGCAANTCGATGACAACTCAAC 5222
 QY 101 SerAlaLeuAspAlaLeuArglyValLeuProThrPheProAspAlaIleuLeuThr 120
 Db 5223 TCGGGCTGGATGCGCTGGCGGTGTCTCTCCCACTTCCCGATGACGCCAATCTTACA 5282
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPAlaLeuThrGlnThrLeuArg 140
 Db 5283 AAGATCGAGACCTGGCTTCCGCCCACTACATCTGGGCACTGACTCAGACGCTGCC 5342
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 Db 5343 ATAGGGACACACAGCTTCTATGCGCGGACCCCTGTCCTCC---TGTGAGAGCTGGGG 5399
 QY 161 SerPro---GlyGlyProProGlyAspTPGlySerLeuTyrSerProValSerGlnAla 179

```

Db      5400 AGCCCCGAGGTGGCTCCACCGGAGACTGCTTACTCTCCAGTCTCCCAAGG 5459
Qy      180 GlycerleuSerProAlaIalaSerleuGlulGluArgProGlyLeuLeuGlyAlaThrSer 199
Db      5460 GGTAACTGAGCCCAACGCGCTTATTGGAGAAATTCCTCGGCTGAGTGCCCAAGCTCC 5519
Qy      200 SerAlaCylSerProGlySerleuAlaPheSerAspPheLeu 214
Db      5520 CCATCTATCTGCTCCCGGAGACACTGCTGTTCTCAGACTTCTTG 5564

RESULT 13
AC109783/c 138070 bp DNA linear HTG 07-FEB-2002
LOCUS AC109783.1 GI:18581594
DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.
ACCESSION AC109783.1
VERSION AC109783.1 GI:18581594
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 138070)
MCCombe,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
Zutavern,T.
TITLE Mouse Genomic Sequence
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138070)
MCCombe,W.R.
AUTHORS Direct Submission
JOURNAL Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
COMMENT -----
Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
Laboratory
Center code: CSHL
Web site: http://www.cshl.org/genseq
Contact: mcombie@cshl.org
----- Project Information
Project name: RP23-121F10
Clone name: RP23-121F10
Insert size: 173000; agarose-fp
Insert size: 141616; sum-of-ctnigs
Quality coverage: 4.00 in Q20 bases; agarose-fp
Quality coverage: 3.70 in Q20 bases; sum-of-ctnigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 17785: contig of 17785 bp in length
* 17786 17874: gap of unknown length
* 17875 30356: contig of 12482 bp in length
* 30357 30444: gap of unknown length
* 30445 42306: contig of 11862 bp in length
* 42307 42394: gap of unknown length
* 42395 53598: contig of 11204 bp in length
* 53599 53686: gap of unknown length
* 53687 64443: contig of 10669 bp in length
* 64444 74016: gap of unknown length
* 74017 74104: contig of 9573 bp in length
* 74105 83366: gap of unknown length
* 83367 83454: gap of unknown length

```

```

* 83455 92355: contig of 8901 bp in length
* 92356 92443: gap of unknown length
* 92444 100821: contig of 8378 bp in length
* 100822 100909: gap of unknown length
* 100910 107529: contig of 6620 bp in length
* 107530 107617: gap of unknown length
* 107618 114066: contig of 6449 bp in length
* 114067 114154: gap of unknown length
* 114155 118873: contig of 4719 bp in length
* 118874 118961: gap of unknown length
* 118962 123619: contig of 4658 bp in length
* 123620 123707: gap of unknown length
* 123708 128240: contig of 4533 bp in length
* 128241 128328: gap of unknown length
* 128329 132682: contig of 4354 bp in length
* 132683 132770: gap of unknown length
* 132771 136341: contig of 3571 bp in length
* 136342 136429: gap of unknown length
* 136430 138070: contig of 1641 bp in length.
FEATURES
source 1..138070
location/Qualifiers
  1..138070
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /db_xref="taxon:10090"
    /clone="RP23-121F10"
BASE COUNT 35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Alignment Scores:
Pred. No.: 4,82e-30 Length: 138070
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
Gaps: 2
US-09-595-947E-10 (1-214) x AC109783 (1-138070)
Qy      1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db      111220 ATGGCGCTCATTCCTTGATGCGCTCAGCAATCCAGAGTGTCCAGACACACAACT 111161
Qy      21 PheProAlaSerGluAlaPglValThrCysProThrSerAlaProProSerProThr 40
Db      111160 TTTCCCGAGAGCTCGGACACAGAAAGTGTCAATTCACACCCCACTAGCCCACT 111101
Qy      41 ArgThrProGlyAsnGlyAlaGluAlaGluGlyGlyCysArgGlyAlaProArgLys 60
Db      111100 CTCATACCTAGAGACTGCTCCGAGCAGAAAGTGGTACTGCCGAGGAGACTCGAGAA 111041
Qy      61 LeuArgAlaArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db      111040 CTCGCGCCGACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 110981
Qy      81 ArgArgSerArgArgLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db      110980 CGAAGAACCGGCGGACAGAAAGCCATATGCGAGCCATTCGATCACAACCTCAAC 110921
Qy      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Db      110920 TCGGCGCTGAGATGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110861
Qy      121 LysIleGluThrPheLeuArgPheAlaHisAspSerTyrIleThrAlaLeuThrGlnThrLeuArg 140
Db      110860 AAGATCGAGACCTCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 110801
Qy      141 IleAlaAspHisSerLeuTyrAlaLeuGluProAlaProHisCysGlyGluLeuGly 160
Db      110800 ATAAGGACACACACTTCTATAGCCCGGAGCCCTCTGAGCC---TGTGAGAGAGCTGGGG 110744
Qy      161 SerPro---GlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
Db      110743 AGCCCCGAGGTGGCTCCACCGGAGACTGCTTACTCTCCAGTCTCCCAAGG 110684

```



```

Oy      161 serPro--GLyG1yProProGlyAspTPG1SerLeuTyrSerProValSerGlnAla 179
Db      206044 AGCCCGGAGGAGGTGGCTCCACAGGGAGCTAGGGGCTCTATCTACTCCAGTCTCCAGCG 206103
Oy      180 GlySerLeuSerProAlaAlaSerLeuGluGluAlaArgProGlyLeuLeuGlyAlaThrSer 199
Db      206104 GGTAACCTGAGGCCGCCACGGGCTCATTTGGAGGAATTCCTCGCTGACAGAGGCCAGACTCC 206165
Oy      200 SerAlaCyLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db      206164 CCATCTCATCTGCTCTCCCGGAGAGCAGCTGGTTCCTCAGACTTCTTG 206208
RESULT 15
AC011010/c
LOCUS      AC011010      170896 bp      DNA      linear      HTG 16-MAR-2000
DEFINITION Homo sapiens clone RP11-6P16, WORKING DRAFT SEQUENCE, 21 unordered
              pieces.
VERSION      AC011010.4  GI:7107881
KEYWORDS      HTG, HTGS_PHASE1, HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      Birren,B., Linton,L., Nusbaum,C. and Lander,E.
AUTHORS      1 (bases 1 to 170896)
              Homo sapiens, clone RP11-6P16
              Unpublished
              2 (bases 1 to 170896)
REFERENCE      Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
AUTHORS      Baldwin,J., Barna,N., Beckelmyr,R., Boguslavsky,L., Bouckgalter,B.,
              Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
              Cooke,P., Deaveliano,K., Dewar,K., Dominko,M., Donelan,L., Doyle,M.,
              Ferreira,P., Fitzhugh,M., Forrest,C., Funke,R., Gage,D.,
              Galagan,J., Gardyna,S., Grant,G., Hagos,B., Headford,A., Horton,L.,
              Howland,J.C., Johnson,R., Jones,C., Kann,L., Karasik,A., Klein,J.,
              Lehoczy,J., Lien,C., Locke,K., MacDonald,P., Marquis,N.,
              McEwan,P., McGuirk,A., McKernan,K., McDonald,I., Meldrum,J.,
              Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
              Petersen,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
              Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
              Testaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
              Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
COMMENT      Direct Submission
              Submitted (29-SEP-1999) Whitehead Institute/MIT Center for Genome
              Research, 320 Church Street, Cambridge, MA 02141, USA
              On Feb 28, 2000 this sequence version replaced gi:6479051.
              All repeats were identified using RepeatMasker:
              Smit, A.F.A. & Green, P. (1996-1997)
              http://ftp.genome.washington.edu/RM/RepeatMasker.html
              ----- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: MIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
              ----- Project Information
              Center project name: L2916
              Center clone name: 6 P16
              ----- Summary Statistics
              Sequencing vector: M13, M7815, 100% of reads
              Chemistry: Dye-terminator Big Dye, 100% of reads
              Assembly program: Phrap, version 0.960731
              Consensus quality: 114103 bases at least Q40
              Consensus quality: 141555 bases at least Q30
              Consensus quality: 158230 bases at least Q20
              Insert size: 154000, agarose-fp
              Insert size: 168896, sum-of-coverage
              Quality coverage: 3.6 in Q20 bases, agarose-fp
              Quality coverage: 3.3 in Q20 bases, sum-of-coverage
              -----
              * NOTE: This is a 'working draft' sequence. It currently
              * consists of 21 contigs. The true order of the pieces
              * is not known and their order in this sequence record is

```

	* arbitrary. Gaps between the contigs are represented by dashes.
	* runs of N, but the exact sizes of the gaps are unknown.
	* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.
	1 1744: contig of 1744 bp in length
	1745 1844: gap of 100 bp
	* 1845 2954: contig of 1110 bp in length
	* 2955 3054: gap of 100 bp
	* 3055 4839: contig of 1785 bp in length
	* 4840 4939: gap of 100 bp
	* 4940 7461: contig of 2522 bp in length
	* 7462 7561: gap of 100 bp
	* 7562 11912: contig of 4351 bp in length
	* 11913 12012: gap of 100 bp
	* 12013 15127: contig of 3115 bp in length
	* 15128 15227: gap of 100 bp
	* 15228 19176: contig of 3949 bp in length
	* 19177 19276: gap of 100 bp
	* 19277 24938: contig of 5562 bp in length
	* 24939 24938: gap of 100 bp
	* 24939 30632: contig of 5594 bp in length
	* 30633 30732: gap of 100 bp
	* 30733 37821: contig of 7089 bp in length
	* 37822 37921: gap of 100 bp.
	* 37922 44638: contig of 6717 bp in length
	* 44639 44738: gap of 100 bp
	* 44739 49424: contig of 4686 bp in length
	* 49425 49524: gap of 100 bp
	* 49525 55517: contig of 6993 bp in length
	* 55518 55617: gap of 100 bp
	* 55618 56613: contig of 8796 bp in length
	* 56614 65513: gap of 100 bp
	* 65514 75659: contig of 10146 bp in length
	* 75660 75759: gap of 100 bp
	* 75760 86433: contig of 10674 bp in length
	* 86434 86533: gap of 100 bp
	* 86534 98763: contig of 12230 bp in length
	* 98764 98863: gap of 100 bp
	* 98864 109919: contig of 11056 bp in length
	* 109920 110019: gap of 100 bp
	* 110020 125264: contig of 15245 bp in length
	* 125265 125364: gap of 100 bp
	* 125365 141043: contig of 15679 bp in length
	* 141044 141143: gap of 100 bp
	* 141144 170896: contig of 29753 bp in length
FEATURES	Location/Qualifiers
SOURCE	1. 170896
	"organism="Homo sapiens"
	"mol_type="genomic DNA"
	"db_xref="taxon:9606"
	"clone_lib="RP11-6P16"
	"clone_id="RP11-6P16"
	"clone_name="RP11-6P16"
	"clone_size="170896"
	"clone_type="genomic DNA"
	"clone_xref="taxon:9606"
	"clone_yeast="Saccharomyces cerevisiae"
	"clone_zymod="XbaI"
	"clone_zymod_size="170896"
	"clone_zymod_type="genomic DNA"
	"clone_zymod_xref="taxon:9606"
	"clone_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_size="170896"
	"clone_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_type="genomic DNA"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_xref="taxon:9606"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_yeast="Saccharomyces cerevisiae"
	"clone_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_zymod_size="170896"
	"clone_zymod_zymod_zymod_zymod_zymod

```
misc_feature 37922..44638 /note="assembly_fragment"  
misc_feature 44739..49424 /note="assembly_fragment"  
misc_feature 49525..56517 /note="assembly_fragment"  
misc_feature 56618..65413 /note="assembly_fragment"  
misc_feature 65514..75659 /note="assembly_fragment"  
misc_feature 75760..86433 /note="assembly_fragment"  
misc_feature 86534..98763 /note="assembly_fragment"  
misc_feature 98864..109919 /note="assembly_fragment"  
misc_feature 11020..125264 /note="assembly_fragment"  
misc_feature 125365..141043 /note="assembly_fragment"  
misc_feature 141144..170896 /note="assembly_fragment"  
misc_feature 170897..170896 /note="assembly_fragment"  
BASE COUNT 41547 a 41714 c 41241 g 44377 t 2017 others  
ORIGIN
```

```
Alignment Scores:  
Pred. No.: 1.39e-15 Length: 170896  
Score: 523.00 Matches: 99  
Percent Similarity: 98.02% Conservative: 0  
Best Local Similarity: 98.02% Mismatches: 2  
Query Match: 46.41% Indels: 0  
DB: 2 Gaps: 0
```

US-09-595-947E-10 (1-214) x AC011010 (1-170896)

```
QY 114 ProAspAspAlaIleuThrLysIleGluThrLeuArgPheAlaHisAsnTyrIleTrp 133  
DB 30631 CCAGACGACGCGAAGTCACCAAGTCAGACGCTCCGCTCCGCCACAACTACATCTGG 30572  
QY 134 AlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeuTyrAlaLeuGluProProAla 153  
DB 30571 GCGCTGACCTCAACGCTGCGCATAGCGGACCAACAGCTGTACGCGCTGGAGCCGCGCGG 30512  
QY 154 ProHisGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173  
DB 30511 CCGCACTGGGGAGCTGGGAGCCAGCGGCTTCCCGGGGACTGGGGGTCCTCTTAC 30452  
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193  
DB 30451 TCCCAAGTCTCCAGGCTGGAGCCCTGATCCCGCGCTCGTGGAGAGAGACCCGGG 30392  
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySerLeuAlaPheSerAspPhe 213  
DB 30391 CTGCTGGGGGACCTTTTCCGCTGCTTGAAGCCAGGAGCTGCTTCTCAGATTCT 30332  
QY 214 Leu 214  
DB 30331 CTG 30329
```

Search completed: February 2, 2004, 21:05:31
Job time : 3780 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 15:49:38 / Search time 271 Seconds
(without alignments)
2131.660 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MTPPSGAPVQVTRTERS.....ICATSSACISPGSLAFSDPL 214

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 2552756 seqs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+.p2n.model -DEV=xlh
-O=/cgn2.1/USPTO/US0955947/runat.02022004.154933.8036/app.query.fasta_1.391
-DB=N.GeneSeq_19Jun03 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pico -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US0955947@cgn2.1 0 @runat.02022004.154933.8036 -NCPUE=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
N.GeneSeq_19Jun03:*
1: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001.DAT:*
23: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneeq/geneeqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed.

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1105	98.0	5340	21 AAC61089	Human neurogenin 3
2	1105	98.0	5340	24 AAD46871	Human neurogenin 3
3	849	75.3	1491	19 AAV42512	cDNA encoding a no
4	819	72.7	804	19 AAV27050	Mouse neurogenin 3
5	819	72.7	804	21 AAZ51981	Murine neurogenin-3
6	819	72.7	861	22 AAF27266	Mouse neurogenin-3
7	819	72.7	861	25 ABV75970	Mouse transcrip
8	819	72.7	1860	24 AAD46872	Murine neurogenin
9	819	72.7	1861	21 AAC61090	Murine neurogenin
10	819	72.7	5567	22 AAF27254	Mouse atonal homol
11	519	46.1	592	24 ABQ49522	Oligonucleotide fo
12	519	46.1	592	24 ABQ49523	Oligonucleotide fo
13	467	41.4	592	24 ABQ49524	Oligonucleotide fo
14	467	41.4	592	24 ABQ49525	Oligonucleotide fo
15	388.5	34.5	790	22 AAF27264	Chicken atonal hom
16	385.5	34.2	1074	22 AAF27263	Chicken atonal hom
17	378.5	33.6	1268	18 AAT74891	Human neurogenic d
18	378.5	33.6	1268	19 AAV42932	DNA encoding human
19	378.5	33.6	1268	25 ABS56390	Human bHLH family
20	378.5	33.6	1665	24 AAD46888	Human neurogenin 1
21	376.5	33.4	714	24 AAD46889	Human neurogenin 1
22	375.5	33.3	1385	19 AAV27049	Mouse neurogenin 2
23	375.5	33.3	1385	21 AAZ51980	Murine neurogenin-2
24	375.5	33.3	1385	22 AAF27269	Mouse neurogenin-2
25	370.5	32.9	1412	22 AAF27255	Mouse atonal homol
26	370.5	32.9	1412	22 AAF27273	Mouse atonal homol
27	370	32.8	1527	19 AAV27045	Rat neurogenin 1 g
28	370	32.8	1527	21 AAZ51976	Rat neurogenin-1 (
29	365	32.4	1332	19 AAV42938	DNA encoding murin
30	365	32.4	1332	25 ABS56396	Mouse bHLH family
31	365	32.4	1333	18 AAT74894	Mouse neurogenic d
32	359.5	31.9	738	19 AAV27046	Mouse neurogenin 1
33	359.5	31.9	738	21 AAZ51977	Murine neurogenin-1
34	346	30.7	6123	24 AAD46890	Human neurogenin-2
35	322.5	28.6	1312	19 AAV27047	Xenopus neurogenin
36	322.5	28.6	1312	21 AAZ51978	Xenopus X-ngnr-1a
37	304	27.0	1277	19 AAV27048	Xenopus neurogenin
38	304	27.0	1277	21 AAZ51979	Xenopus X-ngnr-1b
39	297.5	26.4	778	24 ABQ16590	Oligonucleotide fo
40	297.5	26.4	778	24 ABQ16591	Oligonucleotide fo
41	248	22.0	778	24 ABQ16592	Oligonucleotide fo
42	248	22.0	778	24 ABQ16593	Oligonucleotide fo
43	241	21.4	2161	23 TBL13239	Drosophila melanog
44	241	21.4	4161	23 TBL13238	Drosophila melanog
45	238.5	21.2	1550	22 AAF27276	Mouse atonal homol

ALIGNMENTS

RESULT 1
AAC61089
ID AAC61089 standard; DNA; 5340 BP.
XX
AC AAC61089;
XX
DT 05-FEB-2001 (first entry)
XX
DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
KW islet cell precursor identification; diabetes mellitus; human; ds.
XX
XX
OS Homo sapiens.
XX
XX
FH Key
CDS Location/Qualifiers
3022..3666

```

FT      /tag= a
FT      /product= "Ngn3"
FT      /note= "Neurogenin 3"
PN      WO200059936-A1.
XX      12-OCT-2000.
XX      28-MAR-2000; 2000WO-US08436.
XX      06-APR-1999; 99US-0128180.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI, 2000-664889/64.
XX      P-PSDB; AAY85617.
XX      Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX      them, useful for diagnosis, prevention and treatment of diabetes
XX      mellitus and to identify individuals at risk of diabetes -
XX      Claim 6, Page 46-48; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX      protein AAY85617. The Ngn3 gene is located at chromosome position
XX      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX      protein sequences, and includes an antibody recognising the Ngn3 protein.
XX      Also included in the invention is a method for identifying an islet cell
XX      precursor, the method involves analysing a cell for the expression of the
XX      Ngn3 gene product, where detection of the product is indicative of an
XX      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX      reagent for detecting (in a subject) a predisposition to a defect in
XX      pancreatic islet cell function or formation associated with a defect in
XX      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX      precursor cells expressing Ngn3, and to alter cellular differentiation in
XX      culture in vivo to produce new beta-cells to treat patients with diabetes
XX      mellitus.
SQ      Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;
Alignment Scores:
Pred. No.: 1.65e-59 Length: 5340
Score: 1105.00 Matches: 211
Percent Similarity: 99.07% Conservative: 1
Best Local Similarity: 98.60% Mismatches: 2
Query Match: 98.05% Indels: 0
DB: 21 Gaps: 0
US-09-595-947E-10 (1-214) x AAC61089 (1-5340)
QY      1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB      3022 ATGAGCGCTCAACCTCCGGGTGGCCCACTGTCACAGTACCCTGAGACGAGCGGTCC 3081
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB      3082 TTCCTCCAGACCTCGGAAGAAGAGACCTGCTCCACGTCGCGCCCGCCAGCCCTCACT 3141
QY      41 ArgThrProGlyAsnGlyAlaGluValGluGluGluGluGluGluGluGluGluGluGlu 60
DB      3142 CCGACACCGGGGGAACCTGCCAGAGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3201
QY      61 LeuAlaGluAlaArgArgGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerIysGln 80
DB      3202 CTCCTGGGACCGGCGCGGGGACGACCGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261
QY      81 ArgArgSerArgArgArgIysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB      3262 CGACGAGGTGGGGAAGAGGAGCAACCGCGGAGCGGACATCGAATGCAACAACCTCAAC 3321
QY      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaIysLeuThr 120

```

```

DB      3322 TCGGCACTGAGCGCCCTCGCGGTCTCTGCCCACTTCCAGACGAGCGAAGCTCAC 3381
QY      121 LysIleGluThrThrLeuArgPheAlaHisAsnTrpIleThrPalaLeuThrGluThrLeuArg 140
DB      3382 AAGATCGAGACGCTGCGCTTGGCCCACTACATCTGAGGCGGCTGACTCAACGCTGGCG 3441
QY      141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB      3442 ATAGCGGACACACAGTTGTACCGCTGAGACCGCGCGCGGCACTGGCGGAGCTGGGC 3501
QY      161 SerProGlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAagly 180
DB      3502 AGCCGAGCGGTTCCTCCCGGGGACTGGGGGCTCCCTCACTCCCGCATCTCCAGGCTGGC 3561
QY      181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB      3562 AGCTGAGTCCCGCGCTGCTGAGGAGGACCCGGGCTGCTGGGGGCCACCTCTTCC 3621
QY      201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB      3622 GCTGCTGAGCCCGAGGAGTCTGGCTTTCTCAGATTTTCTG 3663
RESULT 2
AAD46871
ID      AAD46871 standard; DNA; 5340 BP.
XX      AAD46871;
AC      AAD46871;
XX      27-JAN-2003 (first entry)
XX      Human neurogenin 3 (Ngn3) gene.
XX      Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
XX      type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
XX      islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
XX      gene; ds.
XX      Homo sapiens.
XX      Key Location/Qualifiers
FH      CDS 3022..3666
FT      /tag= a
FT      /product= "Human Ngn3 protein"
PN      WO200274045-A2.
XX      26-SEP-2002.
XX      20-MAR-2002; 2002WO-US11166.
XX      20-MAR-2001; 2001US-0817360.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI, 2002-759853/82.
XX      P-PSDB; AAE29277.
XX      Producing a mammalian islet cell for treating diabetes mellitus
XX      comprises introducing into a mammalian cell a nucleic acid molecule
XX      encoding neuroendocrine basic helix-loop-helix transcription factor -
XX      Example 2, Page 87-88; 108pp; English.
XX      The invention relates to a method for producing a mammalian islet cell.
XX      The method comprising introducing into a mammalian cell a nucleic acid
XX      molecule encoding an islet transcription factor for expression of the
XX      islet transcription factor in the cell and for production of islet cell
XX      phenotype in the cell. The islet transcription factor is a neuroendocrine
XX      basic helix-loop-helix (bHLH) transcription factor. The method is useful
XX      for treating type 2 diabetes mellitus and for replacing beta cells lost

```

CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 3
CC (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.

XX Sequence 5340 BP, 1215 A, 1500 C, 1514 G, 1111 T, 0 other;

Alignment Scores:

Alignment No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	1.65e-59	5340	211	1	0	0	0
Percent Similarity:	99.07%						
Best Local Similarity:	98.60%						
Query Match:	98.05%						

US-09-595-947E-10 (1-214) x AAD46871 (1-5340)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 3022 ATGACGGCTCAACCTCGGGTGGCCCACTGTCGAAGTACCCTGAGACGGAGCGTCC 3081
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 3082 TTCCCGAGGCTCGGAGACCAAGTACCTGCTCCCACTGCGCCCGCCAGCCCACT 3141
QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 3142 CGCACACGGGGGAACTGCCACAGCGGAGAGGAGGCTGCCGAGGGCCCCCGAGGAAG 3201
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 3202 CTCGGGCGACGGCGGGGAGAGCGACGCCGCTAAGAGCGAATGGTCATGAGCAAGCAG 3261
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 3262 CCACGAGCGCGGAGAAAGGCGCAACGACCGGAGCGGATCGAATGCAACCTCAAC 3321
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 3322 TCGGCACTGAGCGCTGCGCGGTGCTCTGCCCACTTCCAGACGACCGAGACTCACC 3381
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 3382 AAGATCGAGACGCTGGCTTCCGCCACCACTACATCTGGCGGTGACTCAAGCGTGGCC 3441
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 3442 ATAGCGGACACAGCTTGTACCGCGCTGAGCGCGCCGCCACCTGCGGGAGCTGGGC 3501
QY 161 SerProGlyGlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAlaGly 180
DB 3502 AGCCGAGGCGGTTCCTCCCGGGAGCTGGGGGTCTCTTACTCCCACTCCCAAGGCTGGC 3561
QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB 3562 AGCTGAGTCCGCGCGCGCTGCTGAGAGAGCGACCGGGGCTGGGGGCGCACCTCTTCC 3621
QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 3622 GCTGCTTGAAGCCAGGAGCTGCTTCTCAGATTCTCTG 3663

```

RESULT 3

AAV42512 standard; cDNA, 1491 BP.

AAV42512;

05-OCT-1998 (first entry)

CDNA encoding a novel BHLH protein designated RELAX.

Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;

control; gene expression; transcriptional activator; targeting;

protein expression; central nervous system; CNS; treatment;

KW nervous system disorder; ss.

XX Rattus sp.

XX Key

FT CDS Location/Qualifiers

FT CDS 459..1103

FT CDS /tag= a

FT CDS /product= RELAX

FN W09827206-A2.

PD 25-JUN-1998.

PF 19-DEC-1997; 97WO-FR02368.

PR 19-DEC-1996; 96FR-0015651.

PA (RHON) RHONE-POULENC RORER SA.

PI Mallet J, Ravassard P, Icard-Liepkalns C;

DR WPI, 1998-362775/31.

DR P-PSDB; AAM62991.

PS Basic helix-loop-helix polypeptide and related nucleic acid - with

PT transcriptional activity, for targeting expression of genes to

PT central nervous system and treatment of nervous disease

XX Claim 6, Page 20; 28pp, French.

XX The present sequence encodes a basic helix-loop-helix (BHLH) type

CC protein, designated RELAX (Rat Embryonic Longitudinal Axis) protein.

CC The protein is used to control and participate in gene expression,

CC by acting as transcriptional activator, strictly dependent on the

CC presence of an intact E box (CANNTG), particularly for targeting

CC expression of proteins to the central nervous system (CNS). The

CC nucleic acid sequence can be used to treat nervous system disorders,

CC and antisense sequences can be used to control mRNA transcription.

XX Sequence 1491 BP, 307 A, 487 C, 413 G, 284 T, 0 other;

Alignment Scores:

Alignment No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
1	3.5e-44	1491	166	14	33	2	2
Percent Similarity:	849.00						
Best Local Similarity:	83.72%						
Query Match:	77.21%						

US-09-595-947E-10 (1-214) x AAV42512 (1-1491)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 459 ATGACGGCTCAACCTCGGGTGGCCCACTGTCGAAGTACCCTGAGACGGAGCGTCC 518
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 519 TTCCCGAGGCTCGGAGACCAAGTACCTGCTCCCACTGCGCCCGCCAGCCCACTG 578
QY 41 ArgThrProGlyAsnCyAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 579 CTCGTAACCGAGGAGTGTCTCGAAGACAGAGAGGTGACTGCCGAGGAGCATCGAGGAAG 638
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 639 CTCCTGCGCGCGCGGAGGCGCAACAGCGCCCAAGACCGAGTTGCACTGAGCAAGCAG 698
QY 81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 699 CGACGAGCGCGGAGAGAGCGCAACGCGGGAGCGGACGATGACCACTTAC 758
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

```


Key Location/Qualifiers
 CDS 160..804
 /tag= a
 /product= "Murine neurogenin-3 protein"
 WO200009676-A2.
 24-FEB-2000.
 13-AUG-1999; 99WO-US18525.
 14-AUG-1998; 98US-0096630.
 (CALY) CALIFORNIA INST OF TECHNOLOGY.
 Anderson DJ, Lo L;
 WPI; 2000-256250/22.
 P-PSDB; AAY70570.
 Inducing non-neuronal cells to differentiate into neurons and for non-neuronal cells to express a neuronal subtype-specific marker, comprising contacting the non-neuronal cells with a vector containing neurogenin nucleic acid -
 Claim 1, Fig 1J; 76pp; English.

The patent discloses a method for inducing non-neuronal cells (NNC) to differentiate into neurons and for NNCs to express a neuronal subtype-specific marker. Transformed host cells are used as sources of neuronal and other growth factors; in culture for screening compounds that modulate neural differentiation or as sources of recombinantly produced neurogenin and phoxa proteins for use in transplantation. The cells also have a variety of in vivo uses, e.g. for transplantation at sites of neuronal dysfunction e.g. patients with hearing or vision loss due to optical or auditory nerve damage, brain or spinal cord injuries, and neurodegenerative disorders e.g. Alzheimer's disease. The present sequence encodes murine neurogenin-3 (NGN-3), a transcription factor. NNCs differentiate into neurons through the recombinant expression of a transcription factor that induces a core program of neurogenesis. Forced expression of murine NGN3 can elicit expression of at least some neuronal phenotypic markers even in NNCs.

Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Alignment Scores:
 Pred. No.: 1.35e-42 Length: 804
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 21 Gaps: 2

US-09-595-947E-10 (1-214) x AA251981 (1-804)

1 MetTnPrGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 160 ATGGGCGCTTCATCTTGATGATGGCTCACCACCAAGTCTCCAGACACACAACT 219
 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 220 TTTCCCGAGCGCTCGACACAGAGTGCATTCACATTCACACCCCACTGACCCACT 279
 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyCysArgGlyAlaProArgIys 60
 280 CTCATACCTTGAAGGACCTCCAGACGAGAGTGGGTGCTGCCGAGGAGCTCTCGAGGAAG 339
 61 LeuArgAlaArgArgGlyValArgSerArgProIysSerGluLeuAlaLeuSerIysGln 80
 340 CTCGGGCGGCGAGCGGAGGCGCAACAGGCCCAAGAGGAGTGGCACTGACCAACAG 399
 81 ArgArgSerArgArgIysValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100

DB 400 CGAAGACCGGCGAGAGAGGCAATGATCGGAGCGCCCAATCGCATGCAACCTCAAC 459
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrThrPheProAspAlaValLeuThr 120
 DB 460 TCGGGCGTGAATGCGCTCGGGGTCTCTGCTCCACCTTCCGGATGAGCGCAAACTTACA 519
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPAlaLeuThrGlnThrLeuArg 140
 DB 520 AAGATCGAGACCTTCGCTCGCTTCCCAACATTCATCTGGGCACTACCTGACGCTGCC 579
 QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyValLeuGly 160
 DB 580 ATAGCGGACACAGCTTCTATGAGCCCGGAGCCCTGTGAGCC---TGAGAGAGCTGGGG 636
 QY 161 SerPro---GlyGlyProProGlyValAspThrProGlySerLeuTyrSerProValSerGlnAla 179
 DB 637 AGCCCGGAGGTGGTCCACAGGAGGAGGCTCTATCTACTCTCCAGTCTCCCAAGCG 696
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyValThrSer 199
 DB 697 GGTACCTGAGGCCCGACGCGCTCATTTGAGAGAAATTCCTGGCTTCAGGTGCCAGCTCC 756
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 757 CCATCTATCTGCTCCCGGAGAGCACTGGTGTCTCAGACTTCTTG 801
 RESULT 6
 AAF27266
 ID AAF27266 standard; cDNA, 861 BP.
 AC AAF27266;
 DT 24-APR-2001 (first entry)
 DE Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.
 KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 OS Mus musculus.
 PN WO200073764-A2.
 XX 07-DEC-2000.
 PD 01-JUN-2000; 2000WO-US15410.
 PF 01-JUN-1999; 99US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 PA (BAYU) BAYLOR COLLEGE MEDICINE.
 PI Zoghbi HY, Bellien H, Birmingham N, Hassan B, Ben-Arie N;
 DR WPI; 2001-032190/04.
 P-PSDB; AAB60359.
 PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 PS Disclosure; Page -; 142pp; English.
 XX The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,

CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from Genbank.

SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Alignment Scores:

Pred. No.:	1,45e-42	Length:	861
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2
DB:	22	Gaps:	2

US-09-595-947E-10 (1-214) x AAF27266 (1-861)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 160 ATGGGGCCCTCATCCCTTGATGCGCTCACCATCCAGTGTCCCGAGACACACACACT 219
QY 21 PheProAlaGlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 220 TTTCGCGAGCTCCGAGCAGCAAGTGTCACTTCACATTCACCCCACTAGCCCACT 279
QY 41 ArgThrProGlyAsnCyValaGluGluGluGlyGlyCysArgGlyAlaProArgGly 60
DB 280 CTCATACCTAGGAGCTCTCCGAGCAGAAAGTGGGTGACTGCGAGGAGCTCGAGAGG 339
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerGlyGln 80
DB 340 CTCGCGCCCGGAGCGGAGGCGCCAGACAGGCCAGGAGTGGCACTCAGCAACAG 399
QY 81 ArgArgSerArgArgGlyAlaGluAlaGluAlaGluArgGluArgGluArgGluArg 100
DB 400 CGAAGAGCGCGGCGGAGAGGCGCATGATCGGAGCGCATGCGATGCAACACTCAAC 459
QY 101 SerAlaLeuAlaLeuAlaLeuArgGlyValLeuProThrPheProAlaAlaLeuThr 120
DB 460 TCGGCGCTGATGCGGTGCGGTGCTGCGCCACCTTCGCGATGACCGCAAACTTACA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGlnThrLeuArg 140
DB 520 AAGATGAGAACCTTGCGCTTCCGCCACATCATCTGGGCACTGCTCAGACGCTGCC 579
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 580 ATAGCGGAGCAGACAGCTTCTATGCGCCGAGCGCCCTGTCGCC---TGTGAGAGGCTGCGG 636
QY 161 SerPro---GlyGlyProProGluAspTyrPylSerLeuTyrSerProAlaSerGlnAla 179
DB 637 ACCCGCGGAGGTGGCTCCAGACGGGAGCTGGGGCTCTATCTACTCCCACTCTCCCAAGG 696
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluAlaArgProGlyLeuLeuGlyAlaThrSer 199
DB 697 GGTAACTTACGCGCCACGCGCTCTATGAGGAATTCCTCGGCTGAGAGTCCCACTCC 756
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 757 CCATCTATCTGCTCCCGGAGACACTGAGTGTCTCAGACTTCTTG 801

```

RESULT 7
 ABV75970
 ID ABV75970 standard; cDNA; 861 BP.

XX ABV75970;
 AC 11-FEB-2003 (first entry)
 XX
 DT Mouse transcription factor neurogenin 3 cDNA.
 XX
 DE Mouse; transcription factor; neurogenin 3; ngn3; stem cell;
 XX differentiation; beta-cell; insulin; diabetes; hyperglycaemia;
 KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;
 KW gene; ss.
 XX
 OS Mus musculus.
 XX
 FH Key Location/Qualifiers
 FT CDS 160..804
 FT /*tag=a
 FT /product="Mouse ngn3"
 FT
 WO200286107-A2.
 PD 31-OCT-2002.
 XX
 PF 19-APR-2002; 2002MO-EP04362.
 XX
 PR 19-APR-2001; 2001US-284531P.
 XX
 PA (DEVE-) DEVELOPMENTAL BIOLOGISCHE FORSCH.
 PA (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
 XX
 PI Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;
 XX WPI; 2003-075629/07.
 DR
 XX
 PT Differentiating stem cells into insulin-producing cells useful for
 PT treating pancreatic diseases, by culturing stem cells in suitable
 PT medium and activating gene involved in beta-cell differentiation -
 XX
 PS Disclosure; Page 58-59; 62pp; English.

CC The present sequence is that of cDNA encoding the murine
 CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),
 CC a gene which is required for the specification of the early
 CC endocrine precursor in the pancreatic epithelium and which is
 CC down-regulated once endocrine differentiation begins. The invention
 CC provides a claimed method for differentiating stem cells (especially
 CC embryonic, adult or somatic stem cells and primordial germ cells)
 CC into insulin-producing cells. This involves culturing stem cells in
 CC a suitable medium and activating at least one gene involved in
 CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,
 CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the
 CC delivery of the gene into stem cells using a viral delivery
 CC system, or the delivery of a protein product of the gene into stem
 CC cells. The insulin-producing cells can be transplanted into
 CC animals or human for treatment of pancreatic diseases, metabolic
 CC syndrome and metabolic disorders with impaired glucose tolerance
 CC as diabetes, hyperglycaemia and impaired glucose tolerance
 CC (claimed). The cells can also be used to identify compounds which
 CC stimulate beta-cell differentiation, insulin secretion or glucose
 CC responsiveness. Differentiated beta-cells can be used to study the
 CC toxic and other effects of exogenous compounds on beta-cell
 CC function. In an example from the invention, Pax6 cDNA was inserted
 CC into expression vector PACCMV.plpa under the control of the
 CC cytomegalovirus promoter.

SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;

Alignment Scores:

Pred. No.:	1,45e-42	Length:	861
Score:	819.00	Matches:	163
Percent Similarity:	82.33%	Conservative:	14
Best Local Similarity:	75.81%	Mismatches:	36
Query Match:	72.67%	Indels:	2

```

DB: 25 Gaps: 2
US-09-595-947E-10 (1-214) x ABV75970 (1-861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 160 ATGGGCGCTCATCCCTTGGATGCGCTCAACATCAAGTGTCCCGAGACACAAACACT 219
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 220 TTTCCCGAGGCTTCGAGACCAAGTGTCTCACTTCCATTCCACCCGACTTGGCCCACT 279
QY 41 ArgThrProGlyAsnCyseAlaGluValGluGluGlyCysArgGlyAlaProArgLys 60
DB 280 CTCATACCTAGGAGCTGCTCCGAGAGAGAGTGGGTGATGCGGAGGAGCCTCGAGGAG 339
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 340 CTCCCGCGCCGAGCGGAGGCGCAACAGGCCCAAGAGAGAGTGGCACTCGACAAACG 399
QY 81 ArgArgSerArgArgLysValAlaAsnAspArgGluArgAspArgMetHisAspLeuAsn 100
DB 400 CGAAGAGCGCGGCGAAGAGGCAATGATCGGAGCGCAATCGCATGCAACCTCAAC 459
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 460 TCGGCGCTGGAGCGCTGCGGCTGCTCCGACCTTCCCGGATGACCGCAAACTTACA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnYrIleThrAlaLeuThrGlnThrLeuArg 140
DB 520 AAGATCGAGACCTTCGCTTCCGCCCAACTACATCTGGGCACTGACAGCTGCGC 579
QY 141 IleAlaAspHisSerLeuThrValAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 580 ATAGCGGACCAAGCTTCTATAGCCCGGAGCCCTGTGCC--TGTGAGAGCTGGGG 636
QY 161 SerPro--GlyGlyProProGlyAspTrpGlySerLeuYrSerProValSerGlnAla 179
DB 637 AGCCCGGAGGTGGCTCCCAAGGGAGCTGGGGCTCTATCTATCTCCCACTCCCAAGG 696
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 697 GGTACCTGAGCCCAAGGCTCATTTGAGGAATTCCTGCGCTGAGGTCGCCAGCTCC 756
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 757 CCATCTATCTGCTCCCGAGCACTGTGTCTCAGACTTCTTG 801

```

```

PF 20-MAR-2002; 2002WO-US11166.
XX
PR 20-MAR-2001; 2001US-0817360.
XX
PA (REGC) UNIV CALIFORNIA.
XX
PI German MS, Lin J;
XX
DR MPI; 2002-759853/82.
DR P-FSDB; AAE29278.
XX
PT Producing a mammalian islet cell for treating diabetes mellitus
PR comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor
XX
PS Example 3; Page 89-90; 108pp; English.
XX
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is murine neurogenin 3
CC (Ngn3) gene.
XX
SQ Sequence 1860 BP, 397 A; 559 C; 537 G; 367 T; 0 other;
XX
Alignment Scores:
Pred. No.: 3,216-42 Length: 1860
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 24 Gaps: 2
US-09-595-947E-10 (1-214) x AAD46872 (1-1860)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 1093 ATGGGCGCTCATCCCTTGGATGCGCTCAACATCAAGTGTCCCGAGACACAAACACT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 1153 TTTCCCGAGGCTTCGAGACCAAGTGTCTCAATTCATTCACCCCACTAGCCCACT 1212
QY 41 ArgThrProGlyAsnCyseAlaGluValGluGluGlyCysArgGlyAlaProArgLys 60
DB 1213 CTCATACCTAGGAGCTGCTCCGAGACCAAGTGGTGTACTGCGAGGAGACTCGAGGAG 1272
QY 61 LeuArgAlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 1273 CTCCCGCGCCGAGCGGAGGCGCAACAGGCCCAAGAGCGAGTGGCACTCAGCAAAACG 1332
QY 81 ArgArgSerArgArgLysValAlaAsnAspArgGluArgAspArgMetHisAspLeuAsn 100
DB 1333 CGAAGAGCGCGGCGAAGAGGCAATGATCGGAGCGCAATCGCATCACAACCTCAAC 1392
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 1393 TCGGCGCTGGAGCGCTGCGGCTGCTCCGACCTTCCCGGATGACCGCAAACTTACA 1452
QY 121 LysIleGluThrLeuArgPheAlaHisAsnYrIleThrAlaLeuThrGlnThrLeuArg 140
DB 1453 AAGATCGAGACCTTCGCTTCCGCCCAACTACATCTGGGCACTGACAGAGCTGCGC 1512
QY 141 IleAlaAspHisSerLeuThrValAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
DB 1513 ATAGCGGACCAAGCTTCTATAGCCCGGAGGCCCTGTGCC--TGTGAGAGCTGGGG 1569
QY 161 SerPro--GlyGlyProProGlyAspTrpGlySerLeuYrSerProValSerGlnAla 179

```

```

RESULT 8
AAD46872
ID AAD46872 strand: DNA; 1860 BP.
XX
AC AAD46872;
XX
DT 27-JAN-2003 (first entry)
XX
DE Murine neurogenin 3 (Ngn3) gene.
XX
KW Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
XX
OS Mus musculus.
XX
Key Location/Qualifiers
FH 1093..1737
FT /tag= a
FT /product= "Murine Ngn3 protein"
XX
XX MO200274045-A2.
XX
XX 26-SEP-2002.
XX

```

```

|||||
Db 1570 AGCCCCGAGGTGGCTCCCAACGGGGAGCTGGGCTCTATCTACTCCCAAGTCCCAAGCG 1629
QY 160 GYSerLeuSerProAlaIaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 1630 GGTAACTTGAGCCCAAGGCTCTATTGGAGAAATTCCTGGGCTGAGGTGCCCAAGCTTC 1689
QY 200 SerAlaCyLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 1690 CCATCTTATCTGCTCTCCGGAGACACTGGTGTCTCAGACTTCTTG 1734

RESULT 9
AAC61090
ID AAC61090 standard; DNA; 1861 BP.
XX
AC AAC61090;
XX
DT 05-FEB-2001 (first entry)
XX
DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
KW islet cell precursor identification; mouse; ds.
XX
OS Mus musculus.
XX
FH Key Location/Qualifiers
FT CDS 1093..1737
FT /tag= a
FT /product= "Ngn3"
FT /note= "Neurogenin 3"
XX
XX WO200059936-A1.
XX
XX 12-OCT-2000.
XX
XX 28-MAR-2000; 2000WO-US08436.
XX
XX 06-APR-1999; 99US-0128180.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX German MS, Lin J;
XX
XX MPI; 2000-664989/64.
XX
XX P-PSDB; AAY85618.
XX
XX Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX them, useful for diagnosis, prevention and treatment of diabetes
XX mellitus and to identify individuals at risk of diabetes -
XX
XX Claim 18; Page 49-50; 54pp; English.
XX
XX The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX protein AAY85617. The Ngn3 gene is located at chromosome position
XX 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX protein sequences, and includes an antibody recognising the Ngn3 protein.
XX Also included in the invention is a method for identifying an islet cell
XX precursor, the method involves analysing a cell for the expression of the
XX Ngn3 gene product, where detection of the product is indicative of an
XX islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX reagent for detecting (in a subject) a predisposition to a defect in
XX pancreatic islet cell function or formation associated with a defect in
XX Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX precursor cells expressing Ngn3, and to alter cellular differentiation in
XX culture in vivo to produce new beta-cells to treat patients with diabetes
XX mellitus. The present sequence represents the murine Ngn3 genomic DNA
XX sequence.
XX
XX Sequence 1661 BP; 397 A; 560 C; 537 G; 367 T; 0 other;
XX
XX Alignment Scores:
XX Pred. No.: 3.21e-42 Length: 1861

```

```

Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: 21 Gaps: 2
US-09-595-947E-10 (1-214) x AAC61090 (1-1861)
QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
Db 1093 ATGGCGCTCATCTCTTGATGCGCTGCTACACATCCAGATGTCCTCCAGAACACAAACCT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
Db 1153 TTTCCCGAGCTCGGACCAAGAGTCTCAGTTCCATTCCACCCCACTAGCCCCACT 1212
QY 41 ArgThrProGlyAsnCyAlaGluAlaGluGluGlyGlyValArgGlyValProArgLys 60
Db 1213 CTCATACCTTAGGAGCTGCTCCGAGCAGAGTGGTGACTGCCAGGAGACCTCGAGAG 1272
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
Db 1273 CTCGCGCCCAACGCGGAGGCGGACAGGCCCAAGACGAGTTGCACTCAGCAAAACAG 1332
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
Db 1333 CGAAGAGCCGCGGCGAAGAGGCCAATGATCGGAGCGCAATCGCATGCACAACTCAAC 1392
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
Db 1393 TCGGCGCTGAGATCGCTCGGCGGTCTCTGCGCACTTCCGAGATGAGCCAACTTACA 1452
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTrpIleThrAlaLeuThrGlnThrLeuArg 140
Db 1453 AAGATCGAGACCCCTGCGCTTGCGCCCACTACATCACTGGCACTGACTCAGACGCTGGCG 1512
QY 141 IleAlaAspHisSerLeuTrpAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 1513 ATAGCGGACCAAGCTTATAGCCCGGAGCCCTCTGACC---TGTGAGAGACTGGGG 1569
QY 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuTrpSerProValSerGlnAla 179
Db 1570 AGCCCCGAGGTGGCTCCCAACGGGAGCTGGGCTCTATCTACTCCCAAGTCCCAAGCG 1629
QY 180 GYSerLeuSerProAlaIaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
Db 1630 GGTAACTTGAGCCCAAGGCTCTATTGGAGAAATTCCTGGGCTGAGGTGCCCAAGCTTC 1689
QY 200 SerAlaCyLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
Db 1690 CCATCTTATCTGCTCTCCGGAGACACTGGTGTCTCAGACTTCTTG 1734

RESULT 10
AAF27254
ID AAF27254 standard; cDNA; 5567 BP.
XX
XX AAF27254;
XX
XX 24-APR-2001 (first entry)
XX
XX Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.
XX
XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;
XX cellular proliferation; cerebellar granule neuron; gene therapy;
XX mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
XX transgenic animal; ss.
XX
XX Mus musculus.
XX
XX WO200073764-A2.
XX
XX 07-DEC-2000.

```


XX 01-JUN-2000; 2000MO-US15410.
 XX 01-JUN-1999; 99US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 PA Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 PI WPI: 2001-032190/04.
 XX P-PSDB; AAB60350.
 DR Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX Disclosure; Page -; 142pp; English.
 PS The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
 Alignment Scores:
 Pred. No.: 9, 91e-42 Length: 5567
 Score: 819.00 Matches: 163
 Percent Similarity: 82.338 Conservative: 14
 Best Local Similarity: 75.818 Mismatches: 36
 Query Match: 72.674 Indels: 2
 DB: 22 Gaps: 2
 US-09-595-947E-10 (1-214) x AAF27254 (1-5567)
 QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 DB 4923 ATGGCGCCCTCATCCCTTGATGGCTGACCATCATCAAGTGTCCCAAGACACAACT 4982
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
 DB 4983 TTTCCTCGAGCTTCGAGCAGCAAGTGTCACTTCCAAATTCACCCCACTTGGCCCACT 5042
 QY 41 ArgThrProGlyAsnGlyAlaGluGluGluGluGluGluGluGluGluGluGluGluGlu 60
 DB 5043 CTCATACCTTAGGAGCTCTCCAGAGCAAGTGTGATGCTCCGAGGAGCTTCGAGGAG 5102
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGluValLeuAlaSerArgGln 80
 DB 5103 CTCCTCGCCGACGCGAGGAGGCGCAACAGGCCCAAGAGGAGTGGCACTCAGCAACG 5162
 QY 81 ArgArgSerArgArgGlyGlyValAlaAspArgGluArgAspArgMetHisAspLeuAn 100
 DB 5163 CGAAGAGCGCGGCAAGAGGCGCAATGATGCGAGCGCAATGCAATGCAACCTCAAC 5222
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThr 120

DB 5223 TCGGCGCTGATGCGCTCGGCGGTGTCTGCTCCCACTTCCGAGTACGCCAATTAACA 5282
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTrpIleTPAlaLeuThrGlnThrLeuArg 140
 DB 5283 AAGATCGAGACCTCGCTGCTGCGCAACATCACTTGGGACATGACGCTGCGCC 5342
 QY 141 IleAlaAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGly 160
 DB 5343 ATAGCGACACACAGCTTCTATGCCCCGAGCCCCCTGTGCCC---TGTGAGAGCTGGGG 5399
 QY 161 SerPro---GlyGlyProProGlyAspTrpGlySerLeuThrYSerProValSerGlnAla 179
 DB 5400 AGCCCCGAGGTGGCTCCCAAGGAGACTGGGAGCTTACTACTCCCAAGTCTCCCAAGCG 5459
 QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluAlaProGlyLeuLeuGluAlaThrSer 199
 DB 5460 GGTAACTGAGCCCAAGCCCTCATTTGAGAGAAATTCCTGGGCTTGCAGGTGCCAGCTCC 5519
 QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
 DB 5520 CCATCTATCTGCTCTCCGAGACACTGGTGTCTCAACTTCTTG 5564
 RESULT 11
 ABQ49522/C
 ID ABQ49522 standard, DNA; 592 BP.
 XX ABQ49522;
 XX 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.
 KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 KW SNP; cell differentiation; ds.
 XX Homo sapiens.
 OS WO200218632-A2.
 XX PD 07-MAR-2002.
 XX PD 01-SEP-2001; 2001MO-BP10074.
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX (EPIC-) EPIGENOMICS AG.
 PA Olek A, Piepenbrock C, Berlin K, Guetig D;
 PI WPI: 2002-371829/40.
 DR Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis; comprises selective hybridization of
 PT amplicons from chemically treated DNA -
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.
 PS This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (1) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders

CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

Alignment Scores:

Pred. No.:	4,21e-24	Length:	592
Score:	519.00	Matches:	98
Percent Similarity:	83.67%	Conservative:	25
Best Local Similarity:	66.67%	Mismatches:	24
Query Match:	46.05%	Indels:	0
DB:	24	Gaps:	0

US-09-595-947E-10 (1-214) x ABQ49522 (1-592)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 443 ATAAAGCCTCAACCTCGAATACGCCCTATCTCAATACCCGTAACGAAACGATCC 384
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 Db 389 TTCGCCAAACCTCGAATAACGAAATACCTACCCGACGTCGCCGCCCAACCCCACT 324
 QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgGly 60
 Db 323 CGCACACGAAAAAACTACGCAAAAAAAGAAAAAACTACCGAAAAAAGCCGAAAAA 264
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 263 CTCGCAACGACGCGCAAAACCGCAACCGAATTAACGATTAACCTAAACAAACA 204
 QY 81 ArgArgSerArgArgGlyLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 Db 203 CGACGAAATCGAAGAAAAAACAACGACCGCAACGCAATCGAATACCAACCTCAAC 144
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
 Db 143 TCGACACTTAACGCGCTACGCGATATCCATCCACCTTCCCAACGACGCGAAACTCAC 84
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGlnThrLeuArg 140
 Db 83 AAAATCGAAACGCTACGCTTCGCCCAACCTACATCTTAAACGTAACCTAAACGCTAC 24
 QY 141 IleAlaAspHisSerLeuTyr 147
 Db 23 ATAAAGCAACCAACTTATAC 3
 RESULT 12
 ABQ49523 standard; DNA; 592 BP.
 XX ABQ49523;
 XX 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.
 XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KW gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX Homo sapiens.
 OS
 PN MO200218632-A2.
 XX
 PD 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EPI0074.
 XX 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 XX (EPIG-) EPIGENOMICS AG.
 XX
 XX Olek A, Piepenbrock C, Berlin K, Guetig D,
 DR WPI; 2002-371829/40.
 XX

PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 PT amplicons from chemically treated DNA
 PS
 XX Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridised to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridisation to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridised to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Alignment Scores:	4,21e-24	Length:	592
Pred. No.:	519.00	Matches:	98
Score:	83.67%	Conservative:	25
Percent Similarity:	66.67%	Mismatches:	24
Best Local Similarity:	46.05%	Indels:	0
Query Match:	24	Gaps:	0

US-09-595-947E-10 (1-214) x ABQ49523 (1-592)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
 Db 150 ATAAAGCCTCAACCTCGAATACGCCCTATCTCAATACCCGTAACGAAACGATCC 209
 QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
 Db 210 TTCGCCAAACCTCGAATAACGAAATACCTACCCGACGTCGCCGCCCAACCCCACT 269
 QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgGly 60
 Db 270 CGCACACGAAAAAACTACGCAAAAAAAGAAAAAACTACCGAAAAAAGCCGAAAAA 329
 QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
 Db 330 CTCGCAACGACGCGCAAAACCGCAACCGAATTAACGATTAACCTAAACAAACA 389
 QY 81 ArgArgSerArgArgGlyLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
 Db 390 CGACGAAATCGAAGAAAAAACAACGACCGCAACGCAATCGAATACCAACCTCAAC 449
 QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120

DB 450 TCGACACTAAGCCCTACGCGATTCCTACCCACCCTTCCCAAGACGCGAAGTCAACC 509
 QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTPrAlaLeuThrGlnThrLeuArg 140
 DB 510 AAAATCGAAACCGCTTCGCGCCACACTACATCTTAAACGCTAACCTCAACGCTACGC 569
 QY 141 IleAlaAspHisSerLeuTyr 147
 DB 570 ATTAACGAAACCACTTATAC 590
 RESULT 13
 ABQ49524
 ID ABQ49524 standard; DNA; 592 BP.
 AC ABQ49524;
 XX
 DT 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.
 XX
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPiG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI; 2002-371829/40.
 PT Determining the degree of cytosine methylation in genomic DNA, useful
 PT for diagnosis and prognosis, comprises selective hybridization of
 XX amplicons from chemically treated DNA -
 PS Claim 12; 56pp + Sequence listing; 56pp; German.
 XX
 CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert in a
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridization to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridized to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g., cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.
 XX
 SQ Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;
 Alignment Scores: 7.13e-21 Length: 592
 Pred. No.:

Score: 467.00 Matches: 101
 Percent Similarity: 73.94% Conservative: 4
 Best Local Similarity: 71.13% Mismatches: 37
 Query Match: 41.44% Indels: 0
 DB: 24 Gaps: 0
 US-09-595-947E-10 (1-214) x ABQ49524 (1-592)
 QY 6 SerGIYAlaProThrValGlnValThrArgGluThrArgSerPheProArgAlaSer 25
 DB 165 TCGGTCGCTTATTTGTTTAAGTGAFTCTGAGACGAGCGGTTTTTTTGAAGTTTCG 224
 QY 26 GluAspGluValThrCysProThrSerAlaProProSerProThrArgThrProGlyAsn 45
 DB 225 GAAGACGAAGGATTTGTTTACGTTGCTTTCGTTTATTTTATTCGTATACGGGGGAT 284
 QY 46 CysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProAlaGlyLeuArgAlaArg 65
 DB 285 TGCCTAGAGGCGGAAGAGGAGGTTGTCAGAGGCGTTTCAGAGAGTTTCGGGTACGGCGC 344
 QY 66 GlyGIYArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSerArgArg 85
 DB 345 GGGGACGCTAGTCGCTTTAAGACGAGTTGATGAGTAAAGTACGACGAGGTGGGCA 404
 QY 86 LysIYsAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 105
 DB 405 AAGAACTTAACGATCGGACGCTATCGAATGTATATTTTAATTCGATTTGACGTT 464
 QY 106 LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIleGluThrLeu 125
 DB 465 TTGCGCGGTGTTTGTATTTTATTTTAAAGCATGTTTAAAGTCAAGACGTTG 524
 QY 126 ArgPheAlaHisAsnTyrIleTPrAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
 DB 525 CGTTTCGTTTAATTAATTAATTTTGGCGTTGATTTAAACGTTGCGATTAAGTATAGT 584
 QY 146 LeuTyr 147
 DB 585 TTGTAC 590
 RESULT 14
 ABQ49525/C
 ID ABQ49525 standard; DNA; 592 BP.
 AC ABQ49525;
 XX
 DT 12-JUL-2002 (first entry)
 DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.
 XX
 KM Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 KW drug; side effect; cancer; central nervous system; cardiovascular;
 KM gastrointestinal; respiratory system; single nucleotide polymorphism;
 XX SNP; cell differentiation; ds.
 XX OS Homo sapiens.
 PN WO200218632-A2.
 XX
 PD 07-MAR-2002.
 PF 01-SEP-2001; 2001WO-EP10074.
 XX
 PR 01-SEP-2000; 2000DE-1043826.
 PR 05-SEP-2000; 2000DE-1044543.
 XX
 PA (EPiG-) EPIGENOMICS AG.
 PI Olek A, Piepenbrock C, Berlin K, Guetig D;
 DR WPI; 2002-371829/40.
 PT Determining the degree of cytosine methylation in genomic DNA, useful

PT for diagnosis and prognosis, comprises selective hybridization of
PT amplicons from chemically treated DNA
XX
XX
PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC genomic sample of DNA. The sample is treated chemically to convert
CC cytosine (C) but not methylated C, to uracil, then part of the genomic
CC DNA that contains the target C is amplified to form a labeled amplicon.
CC The amplicon is hybridised to two classes, each with at least one
CC member of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC and the degree of hybridisation to both classes is determined from the
CC label on the amplicon. From the ratio of labels hybridised to the two
CC classes of oligomers, the degree of methylation is calculated. The method
CC is used: (i) for diagnosis and/or prognosis of side effects of
CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC of the central nervous, cardiovascular, gastrointestinal and respiratory
CC systems etc., particularly by detecting mutations or single nucleotide
CC polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
CC types and for investigating cell differentiation. The method allows the
CC methylation status of many C residues to be determined simultaneously.
CC ABQ3410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC method for determining the degree of cytosine methylation described in
CC the disclosure of the invention.

XX
SQ Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Alignment Scores:
Pred. No.: 7,13e-21 Length: 592
Score: 467.00 Matches: 101
Percent Similarity: 73.94% Conservative: 4
Best Local Similarity: 71.13% Mismatches: 37
Query Match: 41.44% Indels: 0
DB: 24 Gaps: 0

US-09-595-947E-10 (1-214) x ABQ49525 (1-592)

QY 6 SerrgialaProthValGlnValThArgGluThrGluArgSerPheProAlaSer 25
DB 428 TCGGGGCGCTTATTGTTTAAGTATTCGTAGACGACGCGTTTATTTTGAAGTTTCG 369
QY 26 GluApgGluValThrCysProThrsSerAlaProPheSerProThrArgThrProGlyAan 45
DB 368 GAGAGCGAAGTATTTGTTTACGTTTCGTTTACGTTTATTCGTTTACGCGGGGAGT 309
QY 46 CysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgGlySerGluAlaArg 65
DB 308 TCCGTAAGAGCGCGAAGAGGAGGTTTCAGGGGTTTCAGAGAACTTTCGGTAACGGCC 249
QY 66 GlyGlyArgSerArgProGlySerGlyGluLeuAlaLeuSerGlyGlnArgArgSerArg 85
DB 248 GGGGGCGTAGTCGGTTTAAAGACGAGTTGATTAAGTAAGACGACGAGTCCGCGCA 189
QY 86 LysLysAlaAspAspArgGluArgGluArgGluArgGluArgGluArgGluArgGlu 105
DB 188 AAGAAAGTTAAGCAGTCGACGAGTATCAATGTAATTAATTAATTCGATTGACGTT 129
QY 106 LeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeu 125
DB 128 TTCCGCGGCTTTGTTTATTTTAAAGACGACGAGTTTATTAAGATCCAGACGTTG 69
QY 126 ArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArgIleAlaAspHisSer 145
DB 68 CGTTTCGTTTATATATATATTTGGCGTTGATTAAACGTTGCGTATACGCGATTATAT 9
QY 146 LeuTyr 147
DB 8 TTGTAC 3

RESULT 15
AAF27264
ID AAF27264 standard; cDNA; 790 BP.

XX
AC AAF27264;
XX
DT 24-APR-2001 (first entry)
XX
DE Chicken atonal homologue ngn2/ath4 cDNA, SEQ ID NO:20.

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;
XX cellular proliferation; cerebellar granule neuron; gene therapy;
XX mechanoreceptive cell growth; auditory; osteopathic; cyostatic;
XX transgenic animal; ss.

XX Gallus gallus.

XX WO200073764-A2.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.

XX 01-JUN-1999; 99US-0137060.

XX 19-JAN-2000; 2000US-0176993.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Zoghbi HY, Belien H, Birmingham N, Hassan B, Ben-Arie N;

XX WPI; 2001-032190/04.

XX P-PSDB; AAB60357.

XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
XX any of its homologs or orthologs, for the treatment of e.g. deafness,
XX osteoarthritis and abnormal cell proliferation -

XX Disclosure; Page -; 142pp; English.

XX The invention relates to the use of atonal-associated nucleic acid or
XX amino acid sequence, or any of its homologues or orthologues as
XX therapeutic agents for the treatment of deafness, partial hearing loss,
XX vestibular effects due to damage or loss of inner hair cells,
XX osteoarthritis and abnormal cell proliferation. The invention also
XX encompasses methods of screening for compounds which affect the
XX expression of an atonal-associated nucleic acid sequence in an animal,
XX and a transgenic animal in which an allele of a native atonal-associated
XX gene is replaced by a heterologous nucleic acid sequence, thus
XX inactivating the atonal-associated allele. The nucleic acids or proteins
XX may be used in a method of treating an animal for hearing impairment,
XX joint disease, balance disorders, abnormal cell proliferation, or other
XX disease related to loss of a functional atonal-associated nucleic acid or
XX protein. They may particularly be used to treat an animal with a
XX deficiency in cerebellar granule neurons or their precursors, and may
XX also be used in promoting mechanoreceptive cell growth and generating
XX hair cells. The present sequence represents an atonal-associated nucleic
XX acid sequence referred to in the invention.
XX Note: The present sequence is not shown in the specification, but
XX was obtained from GenBank.

XX
SQ Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;

Alignment Scores:
Pred. No.: 7,19e-16 Length: 790
Score: 388.50 Matches: 104
Percent Similarity: 54.31% Conservative: 22
Best Local Similarity: 44.83% Mismatches: 63
Query Match: 34.47% Indels: 43
DB: 22 Gaps: 10

US-09-595-947E-10 (1-214) x AAF27264 (1-790)

QY 3 ProGlnProSerGlyAlaProthValGlnValThArgGluThrGluArgSerPhePro 22
DB 47 CCAAGCGCCCGCCGACGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCCGCGCC 103

```
QY 23 Arg-----AlaSerGluAspGluVal 29
Db 104 CGCCCCCGAGATGCGGTGAAGGCGAGAGCCCGCCCGCGCGAGACGACTG 163
QY 30 ThrCysProThrSerAlaProProSerProThrArgThr--ProGlyAsnCysAlaGlu 48
Db 164 CTGCTGCTGCGCTGCGCTGCGCCCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCG 223
QY 49 AlaGluGluGlyGlyCysArgGlyAlaProArgGlyLeuArg-----AlaArgArgGly 66
Db 224 GAGACGAGACGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 283
QY 67 GlyArgSerArgProIysSerGluLeuAlaLeuSerLysGln-----Arg 81
Db 284 GGGCGGCGAGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 343
QY 82 ArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
Db 344 CGAGCGCGGCGGCTGAAAGCCAAACACCGAGCGCGCACCGCATGCACAACTGAACGCG 403
QY 102 AlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLys 121
Db 404 GGGCTGAGACGCGCTGCGCGACGCTGCGCCACCTTCCCGAGGAGCGCCAAAGCTCACAG 463
QY 122 IleGluThrLeuArgPheAlaHisAsnTyrlIeTPAlaLeuThrGlnThrLeuArgIle 141
Db 464 ATCGAGACGCTGCGCTTCCGCCCAACTACATCTGCGGCTCACGAGAGCGCTGCGCTG 523
QY 142 AlaAspHisSer--LeuTyraLeuGluProProAlaProHisCysGlyGluLeuGly 160
Db 524 GCGGCGGCGCGCGCTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 583
QY 161 SerProGlyGlyProProGlyAspTrpGlySerLeuTySerProValSerGlnAlaGly 180
Db 584 AGCCCC--TCGCCCCGCTCGTCTG-----AGCGGC 613
QY 181 SerLeuSerProAlaLaserLeuGluGluArgProGlyLeuLeuGlyValaThrSerSer 200
Db 614 GGGCGCGAGCGCGCGCGCTC-----GCTCGCGCTTAC 646
QY 201 AlaCys--LeuSerProGlySerLeuAlaPheSer 211
Db 647 GCTGCACTTATCGCCGCGAGCGCGCGCGCTCC 682
```

Search completed: February 2, 2004, 20:02:31
Job time : 283 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 18:55:26 ; Search time 355 Seconds
(without alignments)
2197.000 Million cell updates/sec

Title: US-09-595-947E-10

Perfect score: 1127
Sequence: 1 MTPGSGAPTVGVTRETERS.....LGATSSACLSAPSLAFSDPL 214

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2434939 seqs, 1822278265 residues
Total number of hits satisfying chosen parameters: 4869878

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09595947/runat_02022004_154934_8079/app_query.fasta_1.391
-DB=Published Applications NA -QFMT=fastcap -SUFFIX=trpb -MINMATCH=0.1
-LOOPEXT=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62
-TRANS-human40.cdi -LIST=45 -DOCLALIGN=200 -THR SCORE=ppct -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEARSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09595947@cgn_1.1.387@runat_02022004_154934_8079
-NCPU=6 -ICPU=3 -NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPELOCK=100
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FAPOP=6 -FAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELXT=7

Database : Published Applications NA.*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq.*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match length	ID	Description
------------	-------	--------------------	----	-------------

1	1105	98.0	5340	9	US-09-817-360-1	Sequence 1, Appli
2	819	72.7	645	14 <th>US-10-004-717-4</th> <th>Sequence 4, Appli</th>	US-10-004-717-4	Sequence 4, Appli
3	819	72.7	861	14 <th>US-10-004-717-24</th> <th>Sequence 24, Appli</th>	US-10-004-717-24	Sequence 24, Appli
4	819	72.7	1861	9 <th>US-09-817-360-3</th> <th>Sequence 3, Appli</th>	US-09-817-360-3	Sequence 3, Appli
5	388.5	34.5	790	14 <th>US-10-004-717-20</th> <th>Sequence 20, Appli</th>	US-10-004-717-20	Sequence 20, Appli
6	388.5	34.2	1074	14 <th>US-10-004-717-18</th> <th>Sequence 18, Appli</th>	US-10-004-717-18	Sequence 18, Appli
7	375.5	33.3	1385	14 <th>US-10-004-717-30</th> <th>Sequence 30, Appli</th>	US-10-004-717-30	Sequence 30, Appli
8	370.5	33.9	1412	14 <th>US-10-004-717-6</th> <th>Sequence 6, Appli</th>	US-10-004-717-6	Sequence 6, Appli
9	370.5	32.9	1412	14 <th>US-10-004-717-37</th> <th>Sequence 37, Appli</th>	US-10-004-717-37	Sequence 37, Appli
10	370	32.8	1527	8 <th>US-08-722-570-12</th> <th>Sequence 12, Appli</th>	US-08-722-570-12	Sequence 12, Appli
11	359.5	31.9	738	8 <th>US-08-722-570-13</th> <th>Sequence 13, Appli</th>	US-08-722-570-13	Sequence 13, Appli
12	322.5	28.6	1312	8 <th>US-08-722-570-14</th> <th>Sequence 14, Appli</th>	US-08-722-570-14	Sequence 14, Appli
13	304	27.0	1277	8 <th>US-08-722-570-15</th> <th>Sequence 15, Appli</th>	US-08-722-570-15	Sequence 15, Appli
14	238.5	21.2	1550	14 <th>US-10-004-717-13</th> <th>Sequence 43, Appli</th>	US-10-004-717-13	Sequence 43, Appli
15	238.5	21.2	1957	14 <th>US-10-004-717-8</th> <th>Sequence 8, Appli</th>	US-10-004-717-8	Sequence 8, Appli
16	215	19.1	1099	13 <th>US-10-413-358-27</th> <th>Sequence 27, Appli</th>	US-10-413-358-27	Sequence 27, Appli
17	215	19.1	1211	13 <th>US-10-413-358-26</th> <th>Sequence 26, Appli</th>	US-10-413-358-26	Sequence 26, Appli
18	207.5	18.4	993	14 <th>US-10-004-717-47</th> <th>Sequence 47, Appli</th>	US-10-004-717-47	Sequence 47, Appli
19	207.5	18.4	3261	14 <th>US-10-004-717-12</th> <th>Sequence 12, Appli</th>	US-10-004-717-12	Sequence 12, Appli
20	207.5	18.4	3541	14 <th>US-10-004-717-32</th> <th>Sequence 32, Appli</th>	US-10-004-717-32	Sequence 32, Appli
21	206	18.3	1021	12 <th>US-10-321-039-71</th> <th>Sequence 71, Appli</th>	US-10-321-039-71	Sequence 71, Appli
22	205.5	18.2	1056	14 <th>US-10-004-717-10</th> <th>Sequence 10, Appli</th>	US-10-004-717-10	Sequence 10, Appli
23	205.5	18.2	1393	14 <th>US-10-004-717-45</th> <th>Sequence 45, Appli</th>	US-10-004-717-45	Sequence 45, Appli
24	198.5	17.6	675	14 <th>US-10-004-717-15</th> <th>Sequence 15, Appli</th>	US-10-004-717-15	Sequence 15, Appli
25	198.5	17.6	2196	12 <th>US-10-108-260A-1239</th> <th>Sequence 1239, Ap</th>	US-10-108-260A-1239	Sequence 1239, Ap
26	197.5	17.5	485	14 <th>US-10-004-717-59</th> <th>Sequence 59, Appli</th>	US-10-004-717-59	Sequence 59, Appli
27	194	17.2	748	13 <th>US-10-029-386-25014</th> <th>Sequence 25014, A</th>	US-10-029-386-25014	Sequence 25014, A
28	192.5	17.1	1065	14 <th>US-10-004-717-57</th> <th>Sequence 1, Appli</th>	US-10-004-717-57	Sequence 1, Appli
29	192.5	17.1	1572	14 <th>US-10-004-717-57</th> <th>Sequence 57, Appli</th>	US-10-004-717-57	Sequence 57, Appli
30	191.5	17.0	2315	12 <th>US-10-136-728-77</th> <th>Sequence 77, Appli</th>	US-10-136-728-77	Sequence 77, Appli
31	190	16.9	501	14 <th>US-10-004-717-13</th> <th>Sequence 13, Appli</th>	US-10-004-717-13	Sequence 13, Appli
32	187	16.6	948	14 <th>US-10-004-717-41</th> <th>Sequence 41, Appli</th>	US-10-004-717-41	Sequence 41, Appli
33	187	16.6	2993	15 <th>US-10-125-237-56</th> <th>Sequence 56, Appli</th>	US-10-125-237-56	Sequence 56, Appli
34	187	16.6	2993	15 <th>US-10-105-891-56</th> <th>Sequence 56, Appli</th>	US-10-105-891-56	Sequence 56, Appli
35	185.5	16.5	849	13 <th>US-10-226-872-8</th> <th>Sequence 8, Appli</th>	US-10-226-872-8	Sequence 8, Appli
36	183.5	16.3	907	12 <th>US-10-004-717-65</th> <th>Sequence 65, Appli</th>	US-10-004-717-65	Sequence 65, Appli
37	183.5	16.3	932	14 <th>US-10-295-027-1190</th> <th>Sequence 1190, Ap</th>	US-10-295-027-1190	Sequence 1190, Ap
38	183.5	16.3	1791	13 <th>US-10-226-872-1</th> <th>Sequence 1, Appli</th>	US-10-226-872-1	Sequence 1, Appli
39	183.5	16.3	1791	13 <th>US-10-226-872-6</th> <th>Sequence 6, Appli</th>	US-10-226-872-6	Sequence 6, Appli
40	183.5	16.3	1830	13 <th>US-10-226-872-4</th> <th>Sequence 4, Appli</th>	US-10-226-872-4	Sequence 4, Appli
41	183.5	16.3	17290	11 <th>US-09-999-121-7</th> <th>Sequence 7, Appli</th>	US-09-999-121-7	Sequence 7, Appli
42	183.5	16.3	25760	11 <th>US-09-999-121-13</th> <th>Sequence 13, Appli</th>	US-09-999-121-13	Sequence 13, Appli
43	180	16.0	1034	13 <th>US-10-029-386-22762</th> <th>Sequence 22762, A</th>	US-10-029-386-22762	Sequence 22762, A
44	179	15.9	849	13 <th>US-10-226-872-9</th> <th>Sequence 9, Appli</th>	US-10-226-872-9	Sequence 9, Appli
45	177.5	15.7	938	14 <th>US-10-004-717-39</th> <th>Sequence 39, Appli</th>	US-10-004-717-39	Sequence 39, Appli

ALIGNMENTS

RESULT 1
US-09-817-360-1
Sequence 1, Application US/09817360
Patent No. US20020015696A1
GENERAL INFORMATION:
APPLICANT: Lin, Joseph
APPLICANT: German, Michael S.
TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
TITLE OF INVENTION: AND DELIVERY OF INSULIN
FILE REFERENCE: UCSF-129C1P
CURRENT APPLICATION NUMBER: US/09/817.360
CURRENT FILING DATE: 2001-03-20
PRIOR APPLICATION NUMBER: 09/535,145
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: 60/128,180
PRIOR FILING DATE: 1999-04-06
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 5340
TYPE: DNA
ORGANISM: Homo Sapiens
US-09-817-360-1

Alignment Scores:

Pred. No.: 6 88e-92 Length: 5340
 Score: 1105.00 Matches: 211
 Percent Similarity: 99.07% Conservative: 1
 Best Local Similarity: 98.60% Mismatches: 2
 Query Match: 98.05% Indels: 0
 DB: 9 Gaps: 0

US-09-595-947e-10 (1-214) x US-09-817-360-1 (1-5340)

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 3022 ATGAGCGCTCAACCTCGGGTGGCCACTGTCACAGTACCCTGAGACGAGCGGTCC 3081
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 3082 TTCCCGAGGCTCGGAGACGAGTACCTGCCACGCTCCGCGCCCGCCGCGCCACT 3141
QY 41 ArgThrProGlyAsnGlyAlaGluValGluGluGlyCysArgGlyValProArgLys 60
DB 3142 CCGCACCGGGGAACTGCCAGAGGCGGAGAGGAGGCTGCCGAGGGGCCCGAGGAG 3201
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 3202 CTCGGGCGACGGCGCGGGGAGCGACCGGCTTAAGAGCGAGTTGGCACTGAGCAAGCAG 3261
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 3262 CCACGAGATCGCGGAGAAAGGCGACACCGGACCGGACATCGATGACACACCTCAAC 3321
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 3322 TCGGCACTGAGAGCCCTCGCGGCTGCTGCCACCTTCCACAGCAGCGGAGCTCAAC 3381
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIleThrAlaLeuThrGlnThrLeuArg 140
DB 3382 AAGATGAGACGCTGGCTTCCGCCACACTACATCTGGGCGCTGACTCAAGCTGCCG 3441
QY 141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisArgGlyGluLeuGly 160
DB 3442 ATAGCGGACACGACTTGTACGCGCTGAGCGCCGCGCGCGCGCGAGCGAGCTGGGC 3501
QY 161 SerProGlyGlyProProGlyLysPheLysSerLysSerProValSerGlnAlaGly 180
DB 3502 AGCCGAGGCGGTCCCGCGGAGCTGGGGGTCCCTTACTCCCACTTCCAGAGCTGCC 3561
QY 181 SerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSer 200
DB 3562 ACCCTGAGTCCCGCCGCTCGCTGAGAGCGACCCGCGGCTGCTGGGGGCCACTCTCC 3621
QY 201 AlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 3622 GCCTGCTGAGCCCGACGAGCTGGCTTCTCAGATTCTTG 3663

```

RESULT 2
 US-10-004-717-4
 ; Sequence 4, Application US/10004717
 ; Publication No. US2002019265A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGBI, HUDA Y.
 ; APPLICANT: YANG, QI
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; FILE REFERENCE: P01899US4
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; PRIOR FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO: 4
 ; LENGTH: 645
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-004-717-4

US-09-595-947e-10 (1-214) x US-10-004-717-4 (1-645)

Alignment Scores:
 Pred. No.: 1 4e-66 Length: 645
 Score: 819.00 Matches: 163
 Percent Similarity: 82.33% Conservative: 14
 Best Local Similarity: 75.81% Mismatches: 36
 Query Match: 72.67% Indels: 2
 DB: 14 Gaps: 2

```

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 1 ATGAGCGCTCAACCTCGGGTGGCCACTGTCACAGTACCCTGAGACGAGCGGTCC 60
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 61 TTCCCGAGGCTCGGAGACGAGTACCTGCCACGCTCCGCGCCCGCCGCGCCACT 120
QY 41 ArgThrProGlyAsnGlyAlaGluValGluGluGlyCysArgGlyValProArgLys 60
DB 41 CTCATACCTTACGAGCTCTCCAGACGAGTGGTGTACTGCGGAGGAGCTCGAGGAG 180
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 181 CTCGGCGCGGACGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 240
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 241 CCAGAGAGCGGCGGAGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 300
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 301 TCGGCGCTGAGAGCCCTCGCGGCTGCTGCCACCTTCCGAGTACGAGCGGAGCTTACA 360
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrlIleThrAlaLeuThrGlnThrLeuArg 140
DB 361 AAGATGAGACCTCTGCGGCTTCCGCCACACTACATCTGGGACCTGAGCTGAGCGGC 420
QY 141 IleAlaAspHisSerLeuTyrlAlaLeuGluProProAlaProHisArgGlyGluLeuGly 160
DB 421 ATAGCGGACACGACTTGTATGCGCGGAGCGCCCTGTGCGCC--TGTGAGAGCTGGGG 477
QY 161 SerPro---GlyGlyProProGlyLysPheLysSerLysSerProValSerGlnAla 179
DB 478 AGCCCGGAGGTGGCTCCAGCGGAGCTGGGGGTCTTACTTCTCCCGACTTCCAGGCG 537
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 538 GGTAACTGAGCCCAAGCGCTCATTTGAGAGATTCCTCGGCTCGAGGTGCCAGCTCC 597
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 598 CATCTCATCTCTCCCGGAGACACTGGTGTCTCAGACTTCTTG 642

```

RESULT 3
 US-10-004-717-24
 ; Sequence 24, Application US/10004717
 ; Publication No. US2002019265A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGBI, HUDA Y.
 ; APPLICANT: YANG, QI
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; FILE REFERENCE: P01899US4

;; CURRENT APPLICATION NUMBER: US/10/004,717
;; CURRENT FILING DATE: 2002-08-16
;; PRIOR APPLICATION NUMBER: 09/585,645
;; PRIOR FILING DATE: 2000-06-01
;; PRIOR APPLICATION NUMBER: 60/176,993
;; PRIOR FILING DATE: 2000-01-19
;; PRIOR APPLICATION NUMBER: 60/137,060
;; PRIOR FILING DATE: 1999-06-01
;; NUMBER OF SEQ ID NOS: 69
;; SOFTWARE: Patentin Ver. 2.1
;; SEQ ID NO 24
;; LENGTH: 861
;; TYPE: DNA
;; ORGANISM: Mus musculus
US-10-004-717-24

Alignment Scores:
Pred. No.: 1,9e-66 Length: 861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: Gaps: 2

US-09-595-947E-10 (1-214) x US-10-004-717-24 (1-861)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 160 ATGGGCGCTCATCCCTTGATGCGCTCATCATCAAGTGTCTCCAGAGCACAAACACT 219
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 220 TTTCCCGAGGCTTCGAGCACCAAGTGTCTCATTCATTCACCCCACTTGACCTTACT 279
QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 280 CTCATATCTAGGAGCTGCTCCGAGCAGAAAGTGGGTGACTGCGAGGAGCTCGAGGAAG 339
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 340 CTCGCGCGCCGAGCGGAGGCGGCAACAGGCCCAAGAGGAGTGGCACTCGACAAACG 399
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 400 CGAAGAACCGCGCGCAAGAGCCCAATGATCGGAGCGCAATCGCATGCAACCTCAAC 459
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 460 TCGGCGCTGATGCGCTGCGCGGTGCTGCGCACTTCCGAGTGAAGCCAAACTTACA 519
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 520 AAGATCGAACCCCTGCGCTTCCGCAACATCATCTGGGCACTGACGCTGCGCC 579
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyLeuGluGly 160
DB 580 ATAGCGACCAACAGCTTCTATGCGCCGAGCCCTGTCGCC--TGTGAAGAGCTGGGG 636
QY 161 SerPro--GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
DB 637 AGCCCGGAGGTGGCTCCAAAGGGGAGCTGGGCTGATCTACTCCCACTCCCAAGG 696
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 697 GGTAACTGAGCCCAAGCGCTCATTTGAGAGAAATTCCTCGGCTGAGGTGCCAGCTCC 756
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB 757 CCATCTATCTGCTCCGAGAGCACTGCGTGTCTGACACTTCTTG 801

RESULT 4
US-09-817-360-3
; Sequence 3, Application US/09817360

;; Patent No. US20020015696A1
;; GENERAL INFORMATION:
;; APPLICANT: German, Michael S.
;; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
;; TITLE OF INVENTION: AND DELIVERY OF INSULIN
;; FILE REFERENCE: USCF-129CIP
;; CURRENT APPLICATION NUMBER: US/09/817,360
;; CURRENT FILING DATE: 2001-03-20
;; PRIOR APPLICATION NUMBER: 09/535,145
;; PRIOR FILING DATE: 2000-03-24
;; PRIOR APPLICATION NUMBER: 60/128,180
;; PRIOR FILING DATE: 1999-04-06
;; NUMBER OF SEQ ID NOS: 19
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 1861
;; TYPE: DNA
;; ORGANISM: Mus musculus
US-09-817-360-3

Alignment Scores:
Pred. No.: 4,33e-66 Length: 1861
Score: 819.00 Matches: 163
Percent Similarity: 82.33% Conservative: 14
Best Local Similarity: 75.81% Mismatches: 36
Query Match: 72.67% Indels: 2
DB: Gaps: 2

US-09-595-947E-10 (1-214) x US-09-817-360-3 (1-1861)

QY 1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 1093 ATGGGCGCTCATCCCTTGATGCGCTCATCATCAAGTGTCTCCAGAGCACAAACACT 1152
QY 21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThr 40
DB 1153 TTTCCCGAGGCTTCGAGCACCAAGTGTCTCATTCATTCACCCCACTTGACCTTACT 1212
QY 41 ArgThrProGlyAsnCysAlaGluValGluGluGlyGlyCysArgGlyAlaProArgLys 60
DB 1213 CTCATATCTAGGAGCTGCTCCGAGCAGAAAGTGGGTGACTGCGAGGAGCTCGAGGAAG 1272
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGln 80
DB 1273 CTCGCGCGCCGAGCGGAGGCGGCAACAGGCCCAAGAGGAGTGGCACTCGACAAACG 1332
QY 81 ArgArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB 1333 CGAAGAACCGCGCGCAAGAGCCCAATGATCGGAGCGCAATCGCATGCAACCTCAAC 1392
QY 101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB 1393 TCGGCGCTGATGCGCTGCGCGGTGCTGCGCACTTCCGAGTGAAGCCAAACTTACA 1452
QY 121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleThrAlaLeuThrGlnThrLeuArg 140
DB 1453 AAGATCGAACCCCTGCGCTTCCGCAACATCATCTGGGCACTGACGCTGCGCC 1512
QY 141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyLeuGluGly 160
DB 1513 ATAGCGACCAACAGCTTCTATGCGCCGAGCCCTGTCGCC--TGTGAAGAGCTGGGG 1569
QY 161 SerPro--GlyGlyProProGlyAspTrpGlySerLeuTyrSerProValSerGlnAla 179
DB 1570 AGCCCGGAGGTGGCTCCAAAGGGGAGCTGGGCTGATCTACTCCCACTCCCAAGG 1629
QY 180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSer 199
DB 1630 GGTAACTGAGCCCAAGCGCTCATTTGAGAGAAATTCCTCGGCTGAGGTGCCAGCTCC 1689
QY 200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214

Db 507 CGCAGCGTCCTGCGCAGCCTTCCCGCAGCAGCAACCAACTCACCAGAAACCTTGCGC 566
 Qy 127 PheAlaHisSerTyrTleTrrAlaLeuThrGlnThrLeuArgTleAlaAspHisSerLeu 146
 Db 567 TTCGCTTACACTACATACATCGGCGCTTCGCCAGACCTTCGTTGGCCGAGCATGCTTC 626
 Qy 147 -----TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerPro 162
 Db 627 CCTCTCCCGCGCGCTTCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 671
 Qy 163 GlyGlyProProGlyAspTrrGlySerLeuTyrSerPro-----Val 176
 Db 672 GGCACCGCAGCGCGGTTCCGCTGCTCCAGCGGTTCCCGCGCGCGCGCGCGCGCGCG 731
 Qy 177 SerGlnAlaGlySerLeuSerProAlaAlaSer----- 187
 Db 732 TCCGCTCCCGCGCGCGCAGCAGCGCGCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 791
 Qy 188 ---LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla----- 201
 Db 792 GCCCTGCGGCGCTTCGCGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 851
 Qy 202 -----CysLeuSerProGlySerLeuAlaSer 211
 Db 852 TGCCCGTGGTGTCTCCGTCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 887
 RESULT 7
 US-10-004-717-30
 ; Sequence 30, Application US/10004717
 ; Publication No. US2002019265A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGHBI, HUDA Y.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; FILE REFERENCE: P01899USA
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; PRIOR FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 30
 ; LENGTH: 1385
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-004-717-30
 Alignment Scores:
 Pred. No.: 1,84e-25 Length: 1385
 Score: 375.50 Matches: 104
 Percent Similarity: 50.00% Conservative: 19
 Best Local Similarity: 42.28% Mismatches: 55
 Query Match: 33.32% Indels: 69
 DB: 14 Gaps: 10
 US-09-595-947E-10 (1-214) x US-10-004-717-30 (1-1385)
 Qy 26 GluAspGluVal-----ThrCysProThrSerAlaProProSerPro 39
 Db 418 GAAGAGAGAGTACTGATGCTGCTGCGGCTTCGCGCGCTTCGCGCGAGCCTGACCCG 477
 Qy 40 -----ThrArgThrProGlyAsnCys--- 46
 Db 478 ATGCTCTCAGCGCGCAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 537
 Qy 47 -----AlaGluAlaGluGlu-----Gly 52

Db 538 GCGCAGCGTGGAGCGGAAGCCGACAGAGGAGTGGACGAGGAGTCCGCGCTCGGCGCG 597
 Qy 53 GlyCysArgGlyAlaProArgGlyLeuArgAla-----ArgArgGly 66
 Db 598 GGTTCGCGG-----CCAGGCGGCTGTGGCGCTGATGACGAGTGCAGCGTGCAGCGCG 651
 Qy 67 GlyArgSerArg-----ProLysSerGluLeuAlaLeuSerLysGlnArg 81
 Db 652 TCGGCTACCGGCGCGCTTCCTCCGAGTGCACAAAGCGCGAGAGAGAGTGCAGCGATCAG 711
 Qy 82 ArgSerArgArgGlyAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSer 101
 Db 712 AAGACCCGACAGCTCAAGAGCCCAACACCGCGAGCGCAACCCCATGCAACACTTAACGCC 771
 Qy 102 AlaLeuSerAlaLeuArgGlyValLeuProThrPheProAspAspAlaValLeuThrLys 121
 Db 772 GCGCTGACGCGCTGCGCGAGAGTGTCTGCACTTCGCGAGAGTCCAGACTCAGAG 831
 Qy 122 IleGluThrLeuAspPheAlaHisAsnTyrTleTrrAlaLeuThrGlnThrLeuArgTle 141
 Db 832 ATCAGAGCGTGGCTTCGCGCGCAGATTAATGAGGCTCAGCGAGACTGCGCGCTG 891
 Qy 142 AlaAspHisSerLeuTyrAla-----Leu 149
 Db 892 GCGGACCACTGCGCGCGCGCGCGTGGCTCCAGAGGCGCGCTTTCACAGAGCGGCTGCTC 951
 Qy 150 GluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTrr 169
 Db 952 CTGAGCCCGGAGAGTGCCTTCGCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAG 1011
 Qy 170 GlySerLeuTyrSerProValSerGlnAlaGlySerLeuSerPro----- 184
 Db 1012 AGCTGCACCAACAGCGCGCGCTCATCTCCACATTCACGTCCCATACAGCTGCAGCTT 1071
 Qy 185 ---AlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAla----- 201
 Db 1072 TCGCGCGCTGAGCGCGCGTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1130
 Qy 202 -----CysLeuSerPro 205
 Db 1131 TCGTTATGCGGCTCACTT 1148
 RESULT 8
 US-10-004-717-6
 ; Sequence 6, Application US/10004717
 ; Publication No. US2002019265A1
 ; GENERAL INFORMATION:
 ; APPLICANT: ZOGHBI, HUDA Y.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
 ; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
 ; FILE REFERENCE: P01899USA
 ; CURRENT APPLICATION NUMBER: US/10/004,717
 ; PRIOR FILING DATE: 2002-08-16
 ; PRIOR APPLICATION NUMBER: 09/585,645
 ; PRIOR FILING DATE: 2000-06-01
 ; PRIOR APPLICATION NUMBER: 60/176,993
 ; PRIOR FILING DATE: 2000-01-19
 ; PRIOR APPLICATION NUMBER: 60/137,060
 ; NUMBER OF SEQ ID NOS: 69
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 6
 ; LENGTH: 1412
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 US-10-004-717-6
 Alignment Scores:
 Pred. No.: 5.4e-25 Length: 1412
 Score: 370.50 Matches: 106
 Percent Similarity: 48.47% Conservative: 21

Best Local Similarity:	40.46%	Mismatches:	61
Query Match:	32.87%	Indels:	75
DB:	14	Gaps:	11

US-09-595-947E-10 (1-214) X US-10-004-717-6 (1-1412)

[illegible]

```

RESULT 9
US-10-004-717-37
; Sequence 37, Application US/1004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOVAL ASSOCIATED SEQUENCE FOR DAPFNS.
; TITLE OF INVENTION: OSTEOCARTRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054

```

```

: CURRENT APPLICATION NUMBER: US/10/004,717
: CURRENT FILING DATE: 2002-08-16
: PRIOR APPLICATION NUMBER: 09/585,645
: PRIOR FILING DATE: 2000-06-01
: PRIOR APPLICATION NUMBER: 60/176,993
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: 60/137,060
: PRIOR FILING DATE: 1999-06-01
: NUMBER OF SEQ ID NOS: 69
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 37
: LENGTH: 1412
: TYPE: DNA
: ORGANISM: Mus musculus
US-10-004-717-37

```

Alignment Scores:	
Pred. No.:	5.4e-25
Score:	370.50
Percent Similarity:	48.47%
Best local Similarity:	40.46%
Query Match:	32.87%
DB:	14
	Gaps: 11

US-09-595-9475-10 (1-214) x US-10-004-117-37 (1-1412)

Qy 16 GlnThrGluArgSerPheProArgLa-----SerGluAspGluVal 29

Db 73 GAACCCGCGAGATGTTCTGCAAAATCTTAGACTCTGAGCTTTGAAGAGAGAAAGAGAGGTA 133

Qy 30 -----ThrCysProThrSerAlaProPsePro----- 39

Db 133 CTGATGCTGCTGGGCTCCGGCTTCCCGGCTCTGGGAGACCTGACCCCGATGCTCTCAGC 192

Qy 40 -----ThrArgThrProGlyAsnCys----- 46

Db 193 GCGGACGAGAGAGAGACGAGAGAGCTGGCGCGCGGCTCCGGCGCTGGGCGACGCTGA 252

Qy 47 AlaGluAlaGluGlu-----GlyGlyCysArgGly 56

Db 253 GCGGAAGCCGGGAGAGGGGGGTGCAAGGAGCCGCGCTCGGCTGCCGGGGGGTGGCGG-- 308

Qy 57 AlaProArgGlyLeuArgLa-----ArgArgGlyGlyArgSerArg 70

Db 310 ---CCAGGCGGCTGCTGGGCTGATGACGAGAGCAAGCTTCCCGCTCCGCTTACGG 366

Qy 71 -----ProLysSerGluLeuAlaLeuSerIlyGlnArgArgSerArgArg 85

Db 367 GCGGTCCTCCGAGGTCCAGAGCGGGGAGAGCGGTGCGAGCGGCATCAAGAAAGACCCGAGG 428

Qy 86 LysLeuValAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla 108

Db 427 CTCAAGGCCCAACCCGACGAGCGCAACCGCAATGCAACCTAAAGCCGCGCTGAGCGG 486

Qy 106 LeuArgGlyValLeuProThrPheProAspAspAlaValLeuThrIlyValLeuThrLeu 128

Db 487 CTGGCGAGGTGTGCTCCACTTCTCCCGAGGATGCAAGCTTACGAGATGAGAGCGCTG 548

Qy 126 ArgPheAlaHisAsnIlyxIleThrPalaLeuThrGlnThrLeuArgIleAlaAspHisSer 145

Db 547 CGCTTGGCCCAATTAATCACTGGGGGCTCAACGAGACTCTGGCGCTGGCGGACCATGCG 608

Qy 146 LeuIlyrAla-----LeuGluProProAla 153

Db 607 GCGGCGCGCGGTGCTCCAGAGGGGGCGGCTTTACAGGAGGCGGTGCTCTGAGGCCGGGA 666

Qy 154 ProHisLeuGlyGluLeuGlySerProGlyGlyProProGlyAspThrGlySerLeuIlyr 173

Db 667 GCTGCGCTGGCGCCGACGCGGAGACAGCCCTTCTCACCTTCTCTCTGAGCTGCACCAAC 728

Qy 174 SerProValSerIlnIlaGlySerLeuSerPro-----AlaIlaSer 187

Db 727 AGCCCGCGCTCATCTTCAACTTCCAGAGTCCCATTAACAGTGCACCTTATATGCCCGGTAAG 786

QY 188 LeuGIuIaProArGIyLeuGIyAlaThrSerSerAla-----CysLeu 203
DB 787 CC-CGGGTCAGACGCGACTACTGGAGAGCCCGCAGGAGCATCGTTATGCGCC 845
QY 204 SerPro 205
DB 846 TCACCT 851
RESULT 10
US-08-722-570-12
Sequence 12, Application US/08722570
Publication No. US20030044887A1
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Oufu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1527 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-12
Alignment Scores:
Pred. No.: 6.53e-25 Length: 1527
Score: 370.00 Matches: 99
Percent Similarity: 52.97% Conservative: 17
Best Local Similarity: 45.21% Mismatches: 63
Query Match: 32.83% Indels: 40
DB: 8 Gaps: 8
US-09-595-947e-10 (1-214) x US-08-722-570-12 (1-1527)
QY 3 ProGIuIaProArGIyAlaThrValGIuIaThrArgGIuIaThrGluArgSerPhePro 22
DB 293 CCGAGAGCAACAGCGGAGCGGACCTGTCTCAAGTTTCTCAACGACGAGGAGCTGTGCCA 352
QY 23 ArgAlaSerGIu-----AspGIuValThrCysProThr-----SerAlaPro 36
DB 353 GGCTCCAGCGCTTACCTTCACCTGAGGCTGTGCTGCCAGCCGCGAGAGGCGGCCA 412
QY 37 ProSerPro---ThrArgThr---ProGIyAsnCysAlaGluIaGluGIuGIyGIyCysA 55
DB 413 CCTCTCCGGGCGATCGAAGCTTCCCGTGGCCAGGACGAGAGGAGGAGG----- 462

QY 55 rGIyAlaProArGIyLeuGIyAlaArgGIyGIyArGIySerArGIySerGIuIa 75
DB 463 -----CGGCGGCGAGCGGAGGTCGCGCGGCGGTCGCGCCG 502
QY 75 euAlaLeuSerGIyGIuIaArgSerArGIyGIyAlaAsnAspArGIuIaGAsn 95
DB 503 CGTGTCTCAGCTCGCTGCGGAGGAGCGCTCGCTCAAGCCAGATGCGCGGCGCAAC 562
QY 95 rGMeTHiaAspLeuAsnSerAlaLeuAspAlaLeuArgGIyValLeuProThrPhePro 115
DB 563 GTATCATTAACCTCAACGCTGCGGAGGCTGCGGAGGCTGCGGAGGCTGCTGCTGCTCCG 622
QY 115 sPAspAlaLeuThrIyGIyIleGIuIaThrLeuArgPheAlaHisAsnTyrlaPalal 135
DB 623 AGGACCAAGCTCAACGAATTTGAGAGCTGCGCTTGGCTCAACATCACTGCGGCC 682
QY 135 eutHrGIuIaThrLeuArgIleAlaAspHisSerLeu-----TYra 148
DB 683 TGGCTGAGACACTGCGGCTGCGAGATCAAGGAGCTCCGCGGAGCGGTCGCGGAGCGCC 742
QY 148 lAluGIuIaProArGIyAlaProHicGyGIyGIuIaLeuGIySerProGIyIaProProGIyA 168
DB 743 TCCTGCTCCGAGTGTGTCTCCCTGC-----CTGCCGATCCCGAGCCCGGCGAGG 796
QY 168 sp-----TriGIySer-----LeuTyIySerPro 176
DB 797 ATACGAGTCTCGGGGCTCGGGGCGGCTGCTCCCGGCTACTGAGGCTGAGCAGC 856
QY 176 alSerGIuIaGIySerLeuSerProAlaAlaSerLeuGIuIaProGIy 193
DB 857 TCTCTGACCCGAGTATGCTCGGCTTCAGAAAGATTCACCTATGCGCGGCT 909

RESULT 11
US-08-722-570-13
Sequence 13, Application US/08722570
Publication No. US20030044887A1
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Oufu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-13

Alignment Scores:
 Pred. No.: 2,77e-24 Length: 738
 Score: 359.50 Matches: 97
 Percent Similarity: 51.83% Conservative: 16
 Best Local Similarity: 44.50% Mismatches: 66
 Query Match: 31.90% Indels: 40
 DB: 8 Gaps: 8

US-09-595-947E-10 (1-214) x US-08-722-570-13 (1-738)

```

QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPro 22
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 44 CGAGCAGCAACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 23 ArgAlaSer-----GluAspGluValThrCysProThr-----SerAlaPro 36
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 104 GGCTACAGCCCTTACCTCCACCTCGGGGCTCTCTCTGCGACCCGAGAGACGCTCCG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 37 ProSerProThr-----ArgThrProGlyAsnGlyAlaGluGluGlyCys 54
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 164 CCTCTCCGGG--GCATCGAATGTCCTCGGTGCCAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 55 ArgGlyAlaProArgGlyLeuAlaArgArgGlyGlyArgSerArgProLysSerGlu 74
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 214 -----CGGCGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 75 LeuAlaLeuSerLysGlnArgArgSerArgArgLysLysLysLysLysLysLysLys
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 253 GCTCTGCTGACCTCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgLysValLeuProThrPhePro 114
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 313 CGCATGCAACACCTCAACGCTGCGGCTGACGCGCTTGCAGAGGTGCTGCTGCTGCTG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 115 AspAspAlaLysLeuThrLysLysLysLysLysLysLysLysLysLysLysLysLys
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 373 GACGACACCAAGCTCAACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 135 LeuThrGlnThrLeuArgLysLysLysLysLysLysLysLysLysLysLysLysLys
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 433 CTGGCTGAGACACTGCGGCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 148 AlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyLysProGly 167
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 493 CTCTGCTCCCGCAGAGTGTCTCCCTGT-----CTGCGGAGGCGCGGAGCGGCGAC
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 Asp-----TrpGlySer-----LeuLysSerPro 175
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 547 GACACTGAGTCTCGGGGTTCCGGGCGGCTGCTCCCTGCGGCACTGTGCAATCACA
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 176 ValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluGluArgProGly 193
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 607 CTCTGACCCCAAGTAGTCTCTGCGCTTCAAGAACTTCACTATGAGCCCGGAGC 660

```

RESULT 12

US-08-722-570-14 Application US/08722570

Sequence 14, Publication No. US20030044887A1

GENERAL INFORMATION:

APPLICANT: Anderson, David J.

APPLICANT: Ma, Qifu

TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/722,570
 FILING DATE: 27-SEP-1996
 CLASSIFICATION: 5365
 ATTORNEY/AGENT INFORMATION:
 NAME: Silva, Robin M.
 REGISTRATION NUMBER: 38,304
 REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 14:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1312 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 US-08-722-570-14

Alignment Scores:
 Pred. No.: 1.29e-20 Length: 1312
 Score: 322.50 Matches: 86
 Percent Similarity: 54.46% Conservative: 30
 Best Local Similarity: 40.38% Mismatches: 83
 Query Match: 28.62% Indels: 14
 DB: 8 Gaps: 4

US-09-595-947E-10 (1-214) x US-08-722-570-14 (1-1312)

```

QY 3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPro 22
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 380 CGCATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProPro-SerProThrArgTh 42
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 440 TTCCCGGCGAGACGAGAGAGTGCAGCTCGGAGATGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 42 rProGlyAsnCysValaGluAlaGluGluGlyCysArgGlyAlaProArgLysLeuArg 62
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 497 CCGGGGA-----CAGCAGAGGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 62 GAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgR 82
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 542 GAGCCGA-----GGCGGCGCTCAGGGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 82 sSerArgArgLysLysLysLysLysLysLysLysLysLysLysLysLysLysLysLys
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 596 GACCGGCGGCGTTAAAGCTTAACACACCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 102 AlaLeuAspAlaLeuArgLysValLeuProThrPheProAspAlaLysLeuThrLysL 122
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 656 GCTTATTCCTCCAGAGAGAGTGTCTCCCTTACCTGAAGATGCAAACTCACAAGAT
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 122 eGluThrLeuArgPheAlaHisAsnThrLysPheAlaLeuThrGlnThrLeuArgL 142
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 716 AGAGACTTGGCTTGTCTTACACTCACTTGTGGCTCTTGAAGAACTTGGCGCTTGG
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 142 aAspHisSerLeuThrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerP 162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 776 CGACCCAGAGCAGCATGTCTTCAACCCAGACAGACAGACAGCATGTGTGAGAGACTCTTC
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 162 oGlyGlyProProGlyAspTrpGlySerLeuLysSerProValSerGlnAlaGlySerL 182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 836 TTCATCCAGAGAGCCCTCTGAGAGCTGAGCTGCTCTTCTCTCTCTCTCTCTCTCTCT
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 182 UserProAlaAlaSerLeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCy 202
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 896 CTCCCGGCGCAGC-----CTTGCCAGCTCCACCTCGAGACAGATTTGAGTTC 940

```

QY 202 sleuserProGlySerLeuAlaPheSerAapPheLeu 214
DB 941 CTGGCAGCCCTCTGAGCTCCACTGAAACCCCTTCAG 977

RESULT 13
US-08-722-570-15
Sequence 15, Application US/08722570
Publication No. US20030044887A1
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1277 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-15

Alignment Scores:
Pred. No.: 6,29e-19 Length: 1277
Score: 304.00 Matches: 86
Percent Similarity: 57.29% Conservative: 24
Best Local Similarity: 44.79% Mismatches: 64
Query Match: 26.97% Indels: 20
Gaps: 8

US-09-595-947E-10 (1-214) x US-08-722-570-15 (1-1277)

QY 24 AlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThrPro 43
DB 238 AGCTCGAGAGATGAG-----CAGCTACACAGTCCGACCAAGCCCGCTCAGC--- 345

QY 44 GlyAsnCyAlaGluAlaGluGlyCysArgGlyAlaProArgLysLeuArgAla 63
DB 346 -----CACTGACAGCAGGAGCGGACCGAGGAGAGAAACCCCGCATGC 393

QY 64 ArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLysGlnArgArgSer 83
DB 394 AGG-----AGGAGCGGAGCGCGGAGACACCGTCTG---AAGATCAAGAAAGAC 441

QY 84 ArgArgGlyGlyAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeu 103
DB 442 CGGCGCGTTAAAGCCAAATACCGGAGAGAAATCGCATGCACACCTGAATATGCGCTC 501

QY 104 AspAlaLeuArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGlu 123
DB 502 GATTCTCTGAGGAGAGTTCTTACCGCATTAATCCGAGAGCCAAATCAACCAAGATAGAG 561

QY 124 ThrLeuArgPheAlaHisAsnTrpIleTrpAlaLeuThrGlnThrLeuArgIleAlaAsp 143
DB 562 ACCTTGCCCTTGGCCCAACAATACATCTGGGCTCTTACGAAACCTTGCGCTGCGCCGAC 621

QY 144 -----HisLeuLeuTrpAlaLeuLupProProAlaProHisCysGlyGlu-----Leu 159
DB 622 CAGCTGACCGG-ATCTAC---TTCCACCCCGACAGCAGCATATGTTAGACAGACTCTCA 677

QY 160 GlySerProGlyGlyProProGlyAspTrpGlySerLeuTrpSerProValSerGlnAla 179
DB 678 TCCTTCCTCTGA-GCCCTCTCTGAGCTGACGTGCTCCCATCTCCCAACTTTCGACT 736

QY 180 GlySerLeuSerProAlaAlaSerLeuGluArgProGlyLeu---LeuGlyAlaThr 198
DB 737 CCTCTCCCGCCGACGACCTGCGCAGCTCCACCTCGGACAGATATGATGACTGCGAGCCT 796

QY 199 SerSerAlaCysLeuSerProGlySerLeuAlaPhe 210
DB 797 CTGAGCTCCGCTTGAACCCCTTCATGTCGCCCTT 832

RESULT 14
US-10-004-717-43
Sequence 43, Application US/10004717
Publication No. US2002019265A1
GENERAL INFORMATION:
APPLICANT: YANG, OI
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEARNESS,
FILE REFERENCE: P01899054
CURRENT APPLICATION NUMBER: US/10/004,717
CURRENT FILING DATE: 2002-08-16
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: 60/176,993
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: 60/137,060
NUMBER OF SEQ ID NOS: 69
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 43
LENGTH: 1550
TYPE: DNA
ORGANISM: Mus musculus
US-10-004-717-43

Alignment Scores:
Pred. No.: 8,11e-13 Length: 1550
Score: 238.50 Matches: 78
Percent Similarity: 39.29% Conservative: 21
Best Local Similarity: 30.95% Mismatches: 72
Query Match: 21.16% Indels: 81
Gaps: 8

US-09-595-947E-10 (1-214) x US-10-004-717-43 (1-1550)

QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36
DB 106 AGAAAGTTTGCTAGCAATGTGAGACCAAGAAACAAATTAAGAAACAGAGGCTTTCCA 165

QY 37 -----ProSerProThrArgThrProGlyAsnCyAlaGluAla 49
DB 166 AAACAAGTTTCCTTCAGAGAAAGACATTAAGAGCGCCCTCGAGAGAAACCAAGAAA 225

QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59
DB 226 GAAAGAGAGAAAGAAAGACAGAGAGAGAAAGATGAGATGAGC----- 267

```

QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db 268 ---TTGTCAGAGAGGAGGGGCTCAGAGAAAAAGACCACCAAACTGACCTGGAAAG 324
QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db 325 GTC---AAGTTCAGAGAGAGAGAGTAAATGGCGGAGAGAGACCGAGTGCACGGCCCTC 381
QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db 382 AATGATGCTCTGGACAACTTTGGAAAAGTGTCCCTGTTACTCTTAAACCAAAACTG 441
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrTleTyrAlaLeuThrGlnThrLeu 139
Db 442 TCCAAATATGAACCTTACGACTGGCCAAAATTCATCTGGGCACTTTCTGAATTTCTG 501
QY 140 ArgIle----- 141
Db 502 AGGATTGGCAGAGACCGGATCTGCTCAGCTTGTCCAAAACCTTATGCAAGGCTTTCC 561
QY 141 ----- 141
Db 562 CAGCCAACTACAAACTTGTGTGGAGGCTGCTTACAGCTCAGCCAGAGATTCTTGATG 621
QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProProAla 153
Db 622 GGTCAAGGTTGGAGAGGCTGCCACACAGAGTCACTTACTCTTCAATTCACCAACCC 681
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
Db 682 TACCAAGCCCTGAGCTGGCCACTCCCAAGG-----CAT 717
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluLuarProGly 193
Db 718 GGGAGCTCTGATTAATTCAGAGTCCATGAAGAACCTTACATTAATGATGATGATC 777
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db 778 TTCTATGAAGTACCTCCCTCAGTGTGCCAGCCCT 813

RESULT 15
US-10-004-717-8
; Sequence 8, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHEI, HUDA Y.
; APPLICANT: YANG, OI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; PRIOR FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 1957
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-8

```

Alignment Scores:

```

Pred. No.: 1,04e-12 Length: 1957
Score: 238.50 Matches: 78
Percent Similarity: 39.29% Conservative: 21
Best Local Similarity: 30.95% Mismatches: 72
Query Match: 21.16% Indels: 81

```

```

DB: 14 Gaps: 8
US-09-595-947e-10 (1-214) x US-10-004-717-8 (1-1957)
QY 19 ArgSerPheProArgAlaSerGluAsp-----GluValThrCysProThrSerAlaPro 36
Db 192 AAAAAATTGCTTACGAAATGAGAGACCAAGAAACAAATTAAGAAACGAGAGCTTTCCA 241
QY 37 -----ProSerProThrArgThrProGlyAsnCysAlaGluAla 49
Db 242 AAACAAGTTGCTCTCGAGAAAGACATTAAGAGGCCCTCGAGAGAAACGAGAAA 301
QY 50 Glu-----GluGlyGlyCysArgGlyAlaProArg 59
Db 302 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGATGAGATGGC----- 343
QY 60 LysLeuArgAlaArgArgGlyGlyArgSerArgProLysSerGluLeuAlaLeuSerLys 79
Db 344 ---TTGTCAGAGAGAGGGGCTCAGGAAAAAGACCACCAACTGACCTGGAAGAG 400
QY 80 GlnArgSerArgArgLysLysAlaAsnAspArgGluArgAsnArgMetHisAspLeu 99
Db 401 GTC---AAGTTCAGAGAGAGAGAGACTAATGGCGGAGAGAGACCGAGATGCACGGCCCTC 457
QY 100 AsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeu 119
Db 458 AATGATGCTCTGGACAAATTTGGAAAAGTGTCCCTGTTACTCTTAAACCAAAACTG 517
QY 120 ThrLysIleGluThrLeuArgPheAlaHisAsnTyrTleTyrAlaLeuThrGlnThrLeu 139
Db 518 TCCAAATATGAACCTTACGACTGGCCAAAATTCATCTGGGCACTTTCTGAATTTCTG 577
QY 140 ArgIle----- 141
Db 578 AGGATTGGCAGAGACCGGATCTGCTCAGCTTGTCCAAAACCTTATGCAAGGCTTTCC 637
QY 141 ----- 141
Db 638 CAGCCAACTACAAACTGTTGGAGGCTGCTTACAGCTCAAGCCAGAGATTCTTGATG 697
QY 142 -----AlaAspHis-----SerLeuTyrAlaLeuGluProProAla 153
Db 698 GGTCAAGGTTGGAGAGGCTGCCACACAGAGTCACTTACTCTTCAATTCACCAACCC 757
QY 154 ProHisCysGlyGluLeuGlySerProGlyGlyProProGlyAspTyrGlySerLeuTyr 173
Db 758 TACCAAGCCCTGAGCTGGCCACTCCCAAGG-----CAT 793
QY 174 SerProValSerGlnAlaGlySerLeuSerProAlaAlaSerLeuGluLuarProGly 193
Db 794 GGGAGCTCTGATTAATTCAGAGTCCATGAAGAACCTTACATTAATGATGATGATC 853
QY 194 LeuLeuGlyAlaThrSerSerAlaCysLeuSerPro 205
Db 854 TTCTATGAAGTACCTCCCTCAGTGTGCCAGCCCT 889

```

Search completed: February 2, 2004, 21:41:42
 Job time : 361 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: February 2, 2004, 17:09:49 / Search time 2044 Seconds
(without alignments)
2544.598 Million cell updates/sec

Title: US-09-595-947E-10
Perfect score: 1127
Sequence: 1 MPPQSGAPVQVTRTERS.....LGATSSACISFSLAFSDPL 214

Scoring table: BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 22781392 seqs, 12152238056 residues
Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODU=frame+ p2n.model -DEV=xlh
-O=/cgn2.1/USPTO_spool/US0959594/runcat.02022004.154934.8059/app.query.fasta_1.391
-DB=EST -Qfmt=fastap -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LIST=45
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LST=45
-LOCALIGN=200 -THR SCORE=DCT -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFWT=0 -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09595947@cgn2.1.2810@runat.02022004.154934.8059 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-GAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST.*
1: em_estda.*
2: em_estchum.*
3: em_estin.*
4: em_estmu.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_hic.*
9: gb_est1.*
10: gb_est2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estfun.*
16: em_estum.*
17: em_gss_hum.*
18: em_gss_inv.*
19: em_gss_pln.*
20: em_gss_vit.*
21: em_gss_fun.*
22: em_gss_mam.*
23: em_gss_mus.*
24: em_gss_pro.*
25: em_gss_pro.*
26: em_gss_pro.*
27: em_gss_vit.*
28: gb_gss1.*

29: gb_gss2.*
Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	809	71.8	1540	11 AK008017	AK008017 Mus muscu
2	726	64.4	1025	16 BY708009	BY708009 BY708009
3	676.5	60.0	933	28 A2296526	A2296526 RPT-23-1
4	340.5	30.2	592	14 CA979119	CA979119 AGENCOURT
5	327	29.0	865	14 CA96091	CA96091 AGENCOURT
6	327	29.0	1098	13 BQ169355	BQ169355 STR00303
7	318	28.2	603	14 CD282259	CD282259 G39271.16
8	317.5	28.2	632	12 BQ10277	BQ10277 BQ10277
9	314	27.9	687	14 CA945402	CA945402 UI-M-FD0-
10	311	27.6	604	13 BX308104	BX308104 BX308104
11	310.5	27.6	600	12 BG808248	BG808248 2083-52 M
12	299	26.5	600	13 BG924937	BG924937 7103-91 M
13	298.5	26.5	1001	9 AL540071	AL540071 AL540071
14	296.5	26.3	1039	10 BE780690	BE780690 601469349
15	295	26.2	730	13 BU612495	BU612495 UI-M-FR0-
16	285.5	25.3	1037	29 CNS03V19	AL262494 Tetradon
17	281	24.9	823	13 BU054481	BU054481 UI-M-FD0-
18	280	24.8	629	12 BU093114	BU093114 BU093114
19	278	24.7	595	12 BU030202	BU030202 BU030202
20	267.5	23.7	401	9 AW147434	AW147434 da02h12.Y
21	265	23.5	947	13 BX419330	BX419330 BX419330
22	251	22.3	588	9 AV673464	AV673464 AV673464
23	249	22.1	490	9 AV995230	AV995230 AV995230
24	249	22.1	711	13 BW275045	BW275045 BW275045
25	243	21.6	814	13 BQ178789	BQ178789 UI-M-EVO-
26	239	21.2	1003	29 CNS021B1	AL198694 Tetradon
27	234.5	20.8	1038	13 BX419494	BX419494 BX419494
28	231	20.5	1022	13 BU113216	BU113216 603129839
29	229.5	20.4	934	13 BU138911	BU138911 603132642
30	228	20.2	704	13 BU057851	BU057851 UI-M-FR0-
31	228	20.2	740	14 CA319439	CA319439 UI-M-FR0-
32	228	20.2	835	14 CA320553	CA320553 UI-M-FR0-
33	226.5	20.1	710	13 BU057238	BU057238 UI-M-FR0-
34	226.5	20.1	742	13 BQ572268	BQ572268 UI-M-FD0-
35	226.5	20.1	770	13 BQ572426	BQ572426 UI-M-FD0-
36	224.5	19.9	688	10 BG699059	BG699059 602678696
37	224.5	19.9	711	2 HSM067528	Bx478059 Homo sapi
38	224.5	19.9	901	13 BQ424098	BQ424098 AGENCOURT
39	224.5	19.9	1967	11 BC022560	BC022560 Homo sapi
40	222.5	19.7	953	9 AU067624	AU067624 AU067624
41	221.5	19.7	781	13 BU611678	BU611678 UI-M-FR0-
42	221.5	19.7	872	13 BX453565	BX453565 BX453565
43	219.5	19.5	724	10 BE783567	BE783567 601471617
44	219	19.4	722	2 HSM078917	Bx508253 Homo sapi
45	219	19.4	875	29 CNS02BKP	AL189970 Tetradon

ALIGNMENTS

RESULT 1
AK008017
LOCUS
DEFINITION Mus musculus adult male small intestine cDNA, RIKEN full-length (Drosophila), full insert sequence.
ACCESSION AK008017
VERSION AK008017.1 GI:12841941
KEYWORDS HTG; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10349636
2
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11042159
3
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
11076861
4
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
20530913
11076861
5
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
6
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
7
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
8
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
9
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
10
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
11
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
12
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
13
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
14
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
15
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
16
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
17
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
18
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
19
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
20
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
21
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
22
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
23
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
24
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
25
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
26
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
27
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
28
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
29
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
30
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
31
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
32
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
33
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
34
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
35
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
36
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
37
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
38
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
39
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
40
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
41
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
42
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
43
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
44
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
45
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
46
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
47
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
48
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
49
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
50
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
51
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
52
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
53
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
54
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
55
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
56
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
57
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
58
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
59
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
60
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
61
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
62
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
63
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
64
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
65
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
66
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
67
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
68
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
69
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
70
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
71
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
72
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
73
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
74
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
75
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
76
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
77
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
78
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
79
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
80
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
81
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
82
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
83
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
84
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
85
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
86
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
87
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
88
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
89
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
90
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
91
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
92
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
93
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
94
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
95
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
96
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
97
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
98
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
99
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851
100
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
21085660
11217851

COMMENT

Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail: genome-resgsc.riken.go.jp, URL: <http://genome-gsc.riken.go.jp/>, Tel: 81-45-503-9222, Fax: 81-45-503-9216)

Please visit our web site (<http://genome-gsc.riken.go.jp/>) for further details.

FEATURES

source

1. 1540
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:2010001M19"
/db_xref="MGI:1907403"
/db_xref="taxon:10090"
/clone="2010001M19"
/sex="male"
/tissue_type="small intestine"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"

CDS

241..885
/note="unnamed protein product; atonal homolog 5 (Drosophila) [MGD|MG1:893591, GB|NM_009719, evidence: BLASTN, 99%, match=648]
putative"
/codon_start=1
/protein_id="BAB25411.1"
/db_xref="GI:12841942"
/db_xref="MGI:893591"
/translation="MAAPPLDALTQVPEPTQPPGASDHEVSSNSTPPTLIPRDCSEAEVDDCGTSKTLARAGRRRPSSEALSKSRKARDRENNHNSALDALAGVLPPTPDPAKLTKEVTRPAHVYALATQTLADSHLYGPPPVCGGLSGPEGSGNGMGSITSPVQAGNISPASLAEFFGLQVPSSTLPLGALVSDFL"

BASE COUNT

306 a 485 c 397 g 352 t

Alignment Scores:

Pred. No.: 8,47e-46 Length: 1540
Score: 809.00 Matches: 161
Percent Similarity: 82.33% Conservative: 16
Best Local Similarity: 74.88% Mismatches: 36
Query Match: 71.78% Indels: 2
DB: 11 Gaps: 2

US-09-595-947E-10 (1-214) x AK008017 (1-1540)

QY 1 MetTnProGlnProSerGlyValAProThrValGlnValThrArgGluThrArgSer 20
DB 241 ATGGGCGCCATCCCTTGATGCGCTGCACCAACCAAGTGTCCAGAGACACACACT 300
QY 21 PheProArgAlaSerGluAlaSpGluValThrCysProThrSerAlaProProSerProThr 40
DB 301 TTTCGCCGAGCTCCGACACACAGAGTGCCTCAATTCACACCCCACTGACCACT 360
QY 41 AGTThrProGlyAsnCysAlaGluAlaGluGluGlyGlyCysArgGlyAlaProArgGly 60
DB 361 CTCATPACCTTGAAGCTCTCCAGACACAGAGTGGGTGACTGCGAGGAGCCTCGAGGAAG 420
QY 61 LeuArgAlaArgArgGlyGlyArgSerArgProGlySerGluLeuAlaLeuSerArgGln 80

[illegible]

MEDLINE	22354683	
PUBMED	12466851	
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Sphiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9216 Email: genome-res@gs.c.riken.go.jp, URL: http://genome-gsc.riken.go.jp/ Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda .S., Hashizume, M., Hayashida, K., Hirozane, T., Horii, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y., Kondo, S., Konno H., Koye, S., Miyazaki, A., Muraoka, M., Nakamura, M., Nomura, K., Numazaki, R., Ono, M., Oshato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct Submission Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001) Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1611-1630 (2000) RIKEN integrated sequence analysis (RISA) system-384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. Please visit our web site (http://genome-gsc.riken.go.jp) for further details.	
FEATURES	Location/Qualifiers	
source	1. 1025	
	/organism="Mus musculus"	
	/mol_type="mRNA"	
	/db_xref="taxon:10090"	
	/clone="2010001M19"	
	/sex="male"	
	/tissue_type="small intestine"	
	/dev_stage="adult"	
	/lab_host="SOLR"	
	/clone_lib="RIKEN full-length enriched, adult male small intestine"	
	/note="Site 1: XhoI; Site 2: SctI; cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. 1st strand cDNA was primed with a primer [5' GAGGAGAGAGCGCGCCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 5.0 and subtraction to Rot = 20.0. Second strand cDNA was prepared with the primer adapter of sequence [5' GAGGAGAGAGCGCGCCGCAATTATTTCTGAGTAATTAATATCCCCCCCCC 3']. cDNA was cleaved with XhoI and SctI."	
BASE COUNT	215 a 316 c 286 g 207 t	1 others
ORIGIN		
Alignment Scores:	2.75e-40 Length: 1025	
Prod. NO.:	726.00 Matches: 160	
Score:	82.33% Conservative: 17	
Percent Similarity:	74.42% Mismatches: 36	
Best Local Similarity:	64.42% Indels: 7	
Query Match:		

```

DB:      14      Gaps:      2
US-09-595-947E-10 (1-214) x BY708009 (1-1025)
QY      1 MetThrProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB      241 ATGGGGCCCTCATCCCTTGGATGGCTGACCATTCACAGTGTCCCAAGACACAAACCT 300
QY      21 PheProArgAlaSerGluAspGluValThrCysProThrSerAlaProSerProThr 40
DB      301 TTTCCTGGAGCTCTGGACCAACAGAGTCTCAAGTTCATTCACATCCACCTTACCTGACACT 360
QY      41 ArgThrProGlyAsnGlyAlaGluValGluGluGlyCysArgGlyAlaProArgGly 60
DB      361 CTCATACCTAGAGGACTGCTCCGAGCAGAAAGTGGGTGACTGCGGAGGAGCCTCGAGGAG 420
QY      61 LeuArgAlaArgArgGlyGlyValArgSerArgProGlySerGluLeuAlaLeuSerGlyGln 80
DB      421 CTCCTGGCCCGGACGCGGAGGCGGCAACAGGCCCAAGAGGAGTGGCACTCAGCAAAAG 480
QY      81 ArgArgSerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsn 100
DB      481 CAGAGAGCCCGGCGCAGAGAGCCCAATGATCGGAGCGCAATGCAATGCAACCTCAAC 540
QY      101 SerAlaLeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThr 120
DB      541 TCGGGCGCTGGATGCGCTGGCGGTCTCTGCCCACTTCCCGATGACGCCAAACTTACA 600
QY      121 LysIleGluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGluThrLeuArg 140
DB      601 AAGTCGAGACCTGCTGCTCCGCCCAACTATCATCTGGGCACTGACTCAGACGCTGCGC 660
QY      141 IleAlaAspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyGluLeuGly 160
DB      661 ATAGCGGACCAACGCTGTATGAGCCCGGAGCCCCCTGTGCC--TGTGAGAGCTGGGG 717
QY      161 SerProGly---GlyProProGlyAspTyrGlySerLeuTyrSerProValSerGlnAla 179
DB      718 AC-CCCGGACGTGGCTCCCAACGGGAGCTGGGGCTCATATCATCTCCATCTCCCAAGG 776
QY      180 GlySerLeuSerProAlaAlaSerLeuGluGluArgProGlyLeuLeuGluAlaThrSer 199
DB      777 GGTAACCTAGCCCAACG-CCCTCATTTGAGAGATTA-CTTGCGCTGCA-GGTGCACTCC 833
QY      200 SerAlaCysLeuSerProGlySerLeuAlaPheSerAspPheLeu 214
DB      834 CCATCAT-CTGCTCCCGGAGGACTGTGTCTCAGACTTTTG 877

RESULT 3
LOCUS   AZ296526 593 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-160G18.TV RPCI-23 Mus musculus genomic clone RPCI-23-160G18
ACCESSION AZ296526
VERSION   AZ296526.1 GI:9538311
KEYWORDS GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
Zhao, S., Nieman, W., Feldblum, T., Malek, J., Shatsman, S., Akintit,
B., Levine, M., Megam, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
Other GSSs: RPCI-23-160G18.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 0208

```

```

Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pjeter@tigr.org, med.bufileo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.bufileo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_end/mouse/bac_end_intro.html
Plate: 160 row: G column: 18
Seq primer: 17
Class: BAC ends.

FEATURES
    source          location/Qualifiers
    1..593
    /organism="Mus musculus"
    /mol_type="genomic DNA"
    /strain="C57BL/6J"
    /db_xref="taxon:10090"
    /clone="RPCI-23-160G18"
    /sex="Female"
    /lab_host="DH10B"
    /clone_lib="RPCI-23"
    /note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
    EcorI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or
    brain genomic DNA was isolated and partially digested
    with a combination of EcorI and EcorI Methase. Size
    selected DNA was cloned into the pBACe3.6 vector at the
    EcorI sites. The ligation products were transformed into
    DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT      88 a 159 c 213 g 133 t
ORIGIN
Alignment Scores:
Pred. No.:      3,966-37      Length:      593
Score:          676.50      Matches:      136
Percent Similarity: 82.68%      Conservative: 12
Best Local Similarity: 75.98%      Mismatches: 30
Query Match:    60.03%      Indels:      2
DB:             28      Gaps:      1

US-09-595-947E-10 (1-214) x AZ296526 (1-593)
QY      3 ProGlnProSerGlyAlaProThrValGlnValThrArgGluThrGluArgSerPhePro 22
DB      547 CCTCATCCCTTGGATGGCTGCTCCATCCATCCAGTGTCCCAAGACACAAACCTTTTCC 488
QY      23 ArgAlaSerGluAspGluValThrCysProThrSerAlaProProSerProThrArgThr 42
DB      487 GGAGCTCTGGACCAACAGAGTGTCTCAATTCACATTCACCCCACTTACCTCTCATTA 428
QY      43 ProGlyAsnGlyAlaGluValGluGlyCysArgGlyAlaProArgGlySerLeuArg 62
DB      427 CCTAGGAGACTGCTCCGAGCAGAAAGTGGTACTGCCAGGAGACCTCGAGAGAGTCCGC 368
QY      63 AlaArgArgGlyGlyValArgSerArgProLysSerGluLeuAlaLeuSerGlyGlnArg 82
DB      367 GCCCGACGCGGAGGCGCAGACAGGCCCAAGAGCGAGTTGGCACTCAGCAAAACAGGAGA 308
QY      83 SerArgArgGlyValAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAla 102
DB      307 ACCCGCGCAGAAAGCCCAATGATGGAGCGCAATCGCATGCAACCTCACTCGGGG 248
QY      103 LeuAspAlaLeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrLysIle 122
DB      247 CTGGATGGCTGCGGGGTCTCTGCCCACTTCCCGATGACGCCAACTTCAAAAGATC 188
QY      123 GluThrLeuArgPheAlaHisAsnTyrIleTyrAlaLeuThrGluThrLeuArgIleAla 142
DB      187 GAGACCTGCGCTTGGCCCAACATACATCTGGCAGTCACTCAGACGCTGGCAGTAGCG 128
QY      143 AspHisSerLeuTyrAlaLeuGluProProAlaProHisGlyGlyGluLeuGlySerPro 162
DB      127 GACCAAGCTTTATAGGCCCGGAGGCCCTGTGTGCTTG---GAGAGCTGGGAGGCC 72
QY      163 GlyGlyProProGlyAspTyrPlySerLeuTyrSerProValSerGlnAlaGlySer 181

```


/clone="IMAGE:6789795"
 /cissue_type="embryo"
 /lab_host="DH10B (71-resistant)"
 /clone_lib="NCI CGAP Zemb2"
 /note="Vector: pCMV-Sport6.ccdB. Site 1: EcoRV, Site 2:
 NotI; Cloned unidirectionally. Primer: Oligo dt. Average
 insert size 2 kb. Constructed by J. Wang (Research
 Genetics, Invitrogen Corp) from tissue donated by L. Zon
 (Harvard University). Note: this is a NCI CGAP library."

BASE COUNT 226 a 256 c 214 g 169 t
 ORIGIN

Alignment Scores:

Pred. No.: 4,77e-13 Length: 865
 Score: 327.00 Matches: 84
 Percent Similarity: 53.14% Conservative: 26
 Best Local Similarity: 40.58% Mismatches: 65
 Query Match: 29.02% Indels: 32
 DB: 14 Gaps: 6

US-09-595-947e-10 (1-214) x CA496091 (1-865)

QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13
 |||||
 DB 85 CCTCAAGCATCTCCAGCCCAATAGATTATCAAGATGAGATCTGTAATCTCCGAT 144
 |||||
 QY 14 -----ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
 |||||
 DB 145 ATGGAACCTCAAGCTGACCTACTCTTTTGGCAC---ACGATATATGAAGACTCGGCC 201
 |||||
 QY 32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluValGluGlu 51
 |||||
 DB 202 ACAGAGCTCCAGCCCGCTCCCGCGG-----TCTCTCTGC----- 237
 |||||
 QY 52 GlyGlyCysArgGlyAlaProArgGlyLeuArgAlaArgArgGlyValArgSerArgPro 71
 |||||
 DB 238 GGAAAAACCACTGCGCTCCAGCCGCGCTCCAGCAGAAAAAGCGGCGCGCGCG 297
 |||||
 QY 72 LysSerGluLeuAlaLeuSerLysGluArgSerArgArgLysAlaAsnAspArg 91
 |||||
 DB 298 AGGAACGAACCACTGTGACCTGCGTGAAGAAAGACCGAGCTGAAGCCCAACGACCC 357
 |||||
 QY 92 GluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
 |||||
 DB 358 GAGAGGAAACAGATGACCAACCTTAACGACGATTTGATGAGAGCGCTGCTGCT 417
 |||||
 QY 112 ThrPheProAspAspAlaLysLeuThrLysIleGluThrLeuArgPheAlaHisAsnThr 131
 |||||
 DB 418 GCGTTTCTGACGACCAAGCTGACCAAAATTGAGACTGTGCGCTTCTGCTCAACTTAC 477
 |||||
 QY 132 IleThrAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu----- 146
 |||||
 DB 478 ATCTGGGACCTTTCCGAGACCATCCGGATTCGACGACCAAGAGGCGCAAGTCAGAGAC 537
 |||||
 QY 147 ---TyrAlaLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyPro 165
 |||||
 DB 538 GGTCCGCTGCTGCTCCCGAGACTTAAGCTCATGCGACATGACCCGCGGCAAGTAC 597
 |||||
 QY 166 ProGlyAspTyrGlySerLeuThrLysProValSerGlnAlaGlySerLeu----- 182
 |||||
 DB 598 TCTTCTCTGCGCGTGGGGGCGATCTGCTCTTCAACCGCTTACTGCAACTCA 657
 |||||
 QY 183 -----SerProAlaAla 186
 |||||
 DB 658 GACCCGGGACCCCGCGACGC 678

RESULT 6
 B0169355 1098 bp mRNA linear EST 01-MAY-2002
 LOCUS STR00303 segmentation stage cdna library Dario reio cdna clone
 DEFINITION CB260 5' similar to NEUROGENIN 1, mRNA sequence.
 ACCESSION B0169355
 VERSION B0169355.1 GI:20376783

KEYWORDS
 SOURCE
 ORGANISM

EST.
 Dario reio (zebrafish)
 Dario reio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes
 Cyprinidae; Dario.
 1 (bases 1 to 1098)
 Loppin,B., Pflumlo,S., Steffan,T., Heyer,V., Furchauer,M., Thisse
 ,C. and Thisse,B.
 Expression of the zebrafish genome during embryogenesis (2002)

TITLE
 JOURNAL
 COMMENT

Contact: Thisse B
 Institut de Genetique et de Biologie Molculaire et Cellulaire
 CNRS, INSERM, ULP
 1, rue Laurent Fries, BP163, CU de Strasbourg, 67404 Illkirch Cedex
 , France
 Tel: 33 3 88 65 33 60
 Fax: 33 3 88 65 32 01
 Email: thisse@ipmc.u-strasbg.fr
 EST from a cDNA of a gene whose expression is spatially restricted
 during embryogenesis. We have established its expression pattern
 during embryonic development by whole mount in situ hybridization
 on zebrafish embryos from the gastrula stage to 2 days of
 development. The corresponding data are available on the zebrafish
 community database at <http://zfinfo.org/cdna> library preparation: B.
 Riggelman. DNA Sequencing by: IGBMC sequencing facility. Clone
 distribution: zebrafish international resource center at the
 University of Oregon (Institute of Neuroscience, 1254 University of
 Oregon, Eugene, OR 97403-1254)
 Seq primer: TTTATACCTCTCACTAAAGGA.
 Location/Qualifiers
 1..1098
 /organism="Dario reio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="CB260"
 /dev_stage="segmentation stage embryos"
 /clone_libs="segmentation stage cdna library"
 /note="Vector: Lambda zap; Site 1: EcoRI; Site 2: XhoI;
 Oligo dt cDNA library constructed from RNA pooled from
 segmentation stage zebrafish embryos"

FEATURES

source

BASE COUNT 284 a 296 c 264 g 254 t
 ORIGIN

Alignment Scores:

Pred. No.: 6e-13 Length: 1098
 Score: 327.00 Matches: 84
 Percent Similarity: 53.14% Conservative: 26
 Best Local Similarity: 40.58% Mismatches: 65
 Query Match: 29.02% Indels: 32
 DB: 13 Gaps: 6

US-09-595-947e-10 (1-214) x B0169355 (1-1098)

QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13
 |||||
 DB 115 CCTCAAGCATCTCCAGCCCAATAGATTATCAAGATGAGATCTGTAATCTCCGAT 174
 |||||
 QY 14 -----ThrArgGluThrGluArgSerPheProArgAlaSerGluAspGluValThrCys 31
 |||||
 DB 175 ATGGAACCTCAAGCTGACCTACTCTTTTGGCAC---ACGATATATGAAGACTCGGCC 231
 |||||
 QY 32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCysAlaGluValGluGlu 51
 |||||
 DB 232 ACAGAGCTCCAGCCCGCTCCCGCGG-----TCTCTCTGC----- 267
 |||||
 QY 52 GlyGlyCysArgGlyAlaProArgGlyLeuArgAlaArgArgGlyValArgSerArgPro 71
 |||||
 DB 266 GGAAAAACCACTGCGCTCCAGCCGCGCTCCAGCAGAAAAAGCGGCGGCGCGCG 327
 |||||
 QY 72 LysSerGluLeuAlaLeuSerLysGluArgSerArgArgLysAlaAsnAspArg 91
 |||||
 DB 328 AGGAACGAACCACTGTGACCTGCTGAAAGAAAGACCGAGCTGAAGCCCAACGACCC 387
 |||||

QY 52 GluArgAsnArgMetHisAspLeuSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
 |||||
 DB 388 GAGAGAACGATGACCACTTAAACAGCATTCGATCTTTAGAGAGCTCTGCT 447
 |||||
 QY 112 ThrPheProAspAspAlaLeuThrIleuArgGlyValLeuArgPheAlaHisAsnTyr 131
 |||||
 DB 448 GCGTTCTCGACGACCAAGCTGACCAAAATTGAGACTCTGCTGCTCACAACCTAC 507
 |||||
 QY 132 IleTrrAlaLeuThrGlnThrIleuArgGlyValLeuArgPheAlaHisAsnTyr 146
 |||||
 DB 508 ATCTGGGACCTTGGAGACCATCCGATCGACAGACAGAGGAGGAGTCAAGAGAC 567
 |||||
 QY 147 ---TrrAlaLeuGluProProAlaProHisCysGlyIleuGlySerProGlyIlePro 165
 |||||
 DB 568 GGTCCGCTGCTCTCCCGGACTAAGCTGATGAGACGACAGACCCCGGACAGTAC 627
 |||||
 QY 166 ProGlyAspTyrGlySerLeuTyrSerProValSerGlnAlaGlySerLeu----- 182
 |||||
 DB 628 TCTTCTCTCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 687
 |||||
 QY 183 -----SerProAlaAla 186
 |||||
 DB 688 GACCCGCGGACGCCGCGACGC 708
 |||||
 RESULT 7
 LOCUS CD282259 603 bp mRNA linear EST 23-MAY-2003
 DEFINITION G39271.16 NCI CGAP_Zemb2 Danilo, rerio CDNA clone IMAGE:6521499 5',
 mRNA sequence.
 ACCESSION CD282259
 VERSION CD282259.1 GI:31060035
 KEYWORDS EST.
 SOURCE Danilo rerio (zebrafish)
 ORGANISM Danilo rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
 ; Cyprinidae; Danilo.
 1 (bases 1 to 603)
 Amundsen, C., Cachuela, N., Chen, F., Cheung, L. M., Chong, A., Murray, L.,
 Oliveira, J., Park, C., Reyes, J., Yunguen, J., and Swimmer, C.
 Expressed sequence tags from NCI CGAP_Zemb2, a Danilo rerio
 embryonic library
 Unpublished
 Contact: Chen F.
 JOURNAL
 COMMENT Exelixis, Inc.
 170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA
 Tel: 650 837 7000
 Fax: 650 837 8300
 Email: fchen@exelixis.com
 DNA Sequencing by: Exelixis, Inc. Clone distribution information
 can be found through the I.M.A.G.E. Consortium/BLNL at:
 http://image.llnl.gov
 Plate: 14108 row: O column: 3
 High quality sequence stop: 603.
 FEATURES
 location/Qualifiers
 1..603
 /organism="Danilo rerio"
 /mol_type="mRNA"
 /db_xref="taxon:7955"
 /clone="IMAGE:6521499"
 /tissue_type="embryo"
 /lab_host="DH10B (T1-resistant)"
 /clone_1ib="NCI CGAP_Zemb2"
 /note="Vector: PCMV-SPORT6.cdb; Site_1: EcoRV; Site_2:
 NotI; Cloned unidirectionally. Primer: Oligo dT. Average
 insert size 2 kb. Constructed by J. Wang (Research
 Genetics, Invitrogen Corp) from tissue donated by L. Zon
 (Harvard University). Note: this is a NCI CGAP library."
 BASE COUNT 159 a 185 c 155 g 104 t
 ORIGIN
 Alignment Scores:

Pred. No.: 1.39e-12 Length: 603
 Score: 318.00 Matches: 26
 Percent Similarity: 54.89% Conservative: 75
 Best Local Similarity: 41.30% Mismatches: 57
 Query Match: 28.22% Indels: 26
 DB: 14 Gaps: 5
 US-09-595-947E-10 (1-214) x CD282259 (1-603)
 QY 3 ProGlnProSerGlyAlaProThrValGlnVal----- 13
 |||||
 DB 79 CCTCAAGATCTCCAGCCACCAATTAAGATTATCAATGAGAGATGATCTCCGAT 138
 |||||
 QY 14 -----ThrArgGluThrGluArgSerPheProAlaSerGluAspGlyValThrCys 31
 |||||
 DB 139 ATGAAACCTCAAGCTGACTACTCTTTCGCAC---ACGAGATGAGAGACTCGGC 195
 |||||
 QY 32 ProThrSerAlaProProSerProThrArgThrProGlyAsnCyAlaGluAluGlu 51
 |||||
 DB 196 AGCAGCTCCACCCGCGCTCCAGCCGCTCCAGCAGAAAGGCGAGGCGGCGCG 231
 |||||
 QY 52 GlyGlyCysArgGlyAlaProArgIleuArgAlaArgGlyValArgSerArgPro 71
 |||||
 DB 232 GGAAGAACCACTGCGCTCTCCAGCCGCTCCAGCAGAAAGGCGAGGCGGCGCG 291
 |||||
 QY 72 LysSerGluLeuAlaLeuSerLysGlnArgArgSerArgGlyValAlaAspArg 91
 |||||
 DB 292 AGAAGCAAGCACTGTCACGTCGTGAAGAGACCCAGGCTGAGGCGACAGCCGC 351
 |||||
 QY 92 GluArgAsnArgMetHisAspLeuSerAlaLeuAspAlaLeuArgGlyValLeuPro 111
 |||||
 DB 352 GAGAGAACGATGACCACTTAAACAGCATTCGATCTTTAGAGAGCTCTGCT 411
 |||||
 QY 112 ThrPheProAspAspAlaLeuThrIleuArgGlyValLeuArgPheAlaHisAsnTyr 131
 |||||
 DB 412 GCGTTCTCGACGACCAAGCTGACCAAAATTGAGACTCTGCTGCTCACAACCTAC 471
 |||||
 QY 132 IleTrrAlaLeuThrGlnThrIleuArgGlyValLeuArgPheAlaHisAsnTyr 146
 |||||
 DB 472 ATCTGGGACCTTGGAGACCATCCGATCGACAGACAGAGGAGGAGTCAAGAGAC 531
 |||||
 QY 147 ---TrrAlaLeuGluProProAlaProHisCysGlyIleuGlySerProGlyIlePro 165
 |||||
 DB 532 GGTCCGCTGCTCTCCCGGACTAAGCTGATGAGACGACAGACCCCGGACAGTAC 591
 |||||
 QY 166 ProGlyAspTyr 169
 |||||
 DB 592 TCTTCTCTCTG 603
 |||||
 RESULT 8
 LOCUS BU010277 632 bp mRNA linear EST 05-DEC-2001
 DEFINITION BU010277 MF01SSA CDNA Oryzias latipes cdna clone MF01SSA142H01 5',
 mRNA sequence.
 ACCESSION BU010277
 VERSION BU010277.1 GI:17364159
 KEYWORDS EST.
 SOURCE Oryzias latipes (Japanese medaka)
 ORGANISM Oryzias latipes
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 Acanthomorpha; Acanthopterygii; Percomorph; Atherinomorpha;
 Belontiiformes; Adrianchthyidae; Oryziinae; Oryzias.
 1 (bases 1 to 632)
 Kohara, Y., Shin-I, T., Kimura, T., Narita, T., Jindo, T. and Takeda, H.
 Medaka EST Project in Takeda's lab
 Unpublished
 Contact: Tadasi Shin-I
 JOURNAL
 COMMENT Center For Genetic Resource Information
 National Institute of Genetics
 111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855

Email: tshini@genes.nig.ac.jp.

FEATURES
Location/Qualifiers

source

1..632

/organism="Oryzias latipes"

/mol_type="mRNA"

/strain="Hd-IR"

/db_xref="taxon:8090"

/clone="MF01SSA142H01"

/sex="mixture of female and male"

/tissue_type="whole embryo"

/dev_stage="segmentation stage 20 - 25"

/clone_lib="MF01SSA cDNA"

BASE COUNT

120 a 231 c 190 g 90 t

1 others

ORIGIN

Alignment Scores:

Pred. No.: 1.58e-12

Score: 317.50

Percent Similarity: 50.45%

Best Local Similarity: 41.36%

Query Match: 28.17%

DB: 12

Matches: 632

Conservative: 20

Mismatches: 55

Indels: 55

Gaps: 11

US-09-595-947e-10 (1-214) x BJ010277 (1-632)

2 ThrProGln-----ProSerGlyAlaProThrValGlnValThrArgGlnThr

92 ACCCCCGCATTTGGAACCCGCTCCGCTGGCATGATCCGCGATTCGACATCGACA

18 GluArgSer-----PheProArgAlaSerGluAspGluValThrCysProThrSer

152 GC-AAACAGCTGGACATTTTTCACACACTGACGAGAGAGACCCGAGTGGC-----

35 AlaProProSerProThrArgThrProGlyAsnGlyAlaGluAlaGluGlyGlyCys

205 CGCCCACTCTCTCCGAGCGC-----

55 ArgGlyAlaProArgGlyLeuArgAlaArgGlyGlyArgSerArgProGlySerGlu

226 -----GGGCAACAGCAAGAAAGAGAGCGCGCGAGCGCGCGCGCGC-----TGGCAG

75 LeuAlaLeuSerLeuGlnArgSerArgGlyGlyGlyAlaAsnAspArgGluArgan

271 GCGGTCTGCGAGGTGGTGAAGAAAGAGCGCGCTGAAAGCTTAAACGACCGGAGCGAAG

95 ArgMetHisAspLeuAsnSerAlaLeuAspAlaLeuArgGlyValLeuProThrPhePro

331 CGCATGCAATACCTGAACAGCGCGCTGAGTGAAGCTGCGCGCGCTCTGCGCGCTTCCG

115 AspAspAlaLeuLeuThrLysLysLysLysLysLysLysLysLysLysLysLysLys

391 GACGAACCAAGCTGCAAGATCGAGACTCTGCGCTTGGCGACAACTCAATCTGGCT

135 LeuThrGlnThrLeuArgGlnLeuAlaAspHisSerLeuThrAla-----LeuGluPro

451 CTGTCCGAGACCATCGCATCGCAGAC-----CTGCAAGCGGAGGCGGCAACGACAC

152 ProAla-----ProHisCys-----GlyGluLeuGlySerProGlyGlyProGly

505 CTTCTGCTGCTTAACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

168 AspTTPGlySerLeuThrLysLysLysLysLysLysLysLysLysLysLysLysLys

565 AGCTGG-----AGCCCGACGCGATCTCTGCGCTCTGCGCGCGG-----

188 LeuGluGluArgProGlyLeuLeuGlyAlaThrSerSerAlaCysLeuSerProGlySer

607 -----TACTGCGCTGTCAGACCGCGCGGCGAGC

630

RESULT 9

CA945402

LOCUS

DEFINITION

UI-M-FDO-cdh-1-12-0-UI.r1 NIH_BMAP_FDO Mus musculus cDNA clone

IMAGE: 6828925 5', mRNA sequence.

ACCESSION
CA945402
VERSION
CA945402.1 GI:27433882KEYWORDS
EST.SOURCE
Mus musculus (house mouse)ORGANISM
Mus musculusREFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.AUTHORS
1 (bases 1 to 687)TITLE
NIH-MGC http://imgc.nci.nih.gov/JOURNAL
National Institutes of Health, Mammalian Gene Collection (MGC)COMMENT
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMIL at:
http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: PYX-5.

FEATURES
Location/Qualifiers

1..687

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/clone="IMAGE: 6828925"

/tissue_type="whole brain"

/dev_stage="embryo 12.5 dpc"

/lab_host="DH10B (T1 phage resistant)"

/clone_lib="NIH BMAP FDO"

/note="Organ: brain, Vector: PYX-Asc, Site 1: EcoR I;
Site 2: Not I; The library was constructed according to
Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with EcoR I adaptor, digested with Not I, and then
cloned directionally into PYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is TGAAGAGAGC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT

117 a 228 c 250 g 91 t

1 others

ORIGIN

Alignment Scores:

Pred. No.: 2.96e-12

Score: 314.00

Percent Similarity: 54.50%

Best Local Similarity: 43.92%

Query Match: 27.86%

DB: 14

Matches: 687

Conservative: 20

Mismatches: 28

Indels: 58

Gaps: 9

US-09-595-947e-10 (1-214) x CA945402 (1-687)

16 GluThrGluArgSerPheProArgAla-----SerGluAspGluVal

77 GAGCCGCGTAGAGTGTCTCAATCTGAGACTCGAGCTGACAGAGAGAGAGAGAGTA

30 -----ThrcyProThrSerAlaProProSerPro-----

137 CTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG

40 -----ThrcyProThrSerAlaProProSerPro-----

RESULT 9

CA945402

LOCUS

DEFINITION

UI-M-FDO-cdh-1-12-0-UI.r1 NIH_BMAP_FDO Mus musculus cDNA clone

/note=Vector: pRT7JD-pac; Clone distribution : AGENAE
Resource Centre. Francois PIUMI,
Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
Etude du genome (LRBG), Domaine de Vilvert, 78352,
Jouy-en-Josas cedex, "FRANCE"

BASE COUNT 162 a 178 c 148 g 116 t

ORIGIN

Alignment Scores:

Pred. NO.: 4.2e-12 Length: 604

Score: 311.00 Matches: 64

Percent Similarity: 68.60% Conservative: 19

Best Local Similarity: 52.89% Mismatches: 32

Query Match: 27.60% Indels: 6

DB: 13 Gaps: 2

US-09-595-947E-10 (1-214) x BX308104 (1-604)

OY 26 GLUASGGILUALITRYCPProThr-----SerAlaProPProSerProThrArgThr 42

Db 203 GATACAGAAATTGTGCAGCAGCATGCAACCGCTTCCTCCCTTCATCTCCATCGGCAAG 262

OY 43 ProGlyAncySAIagIuaIagLUGIGLYCYsarArgGIalAProArgLyLeuArg 62

Db 263 CCCGCCCTCCGGCGACGACGACGACGAGCGTCTGCTCTACGATCTGGCACAGAAGAG 322

OY 63 AlAargArgGIgIyArGserArgProLySergIuleuAlaleuSerLySGInArgArg 82

Db 323 AGACGCAGAGGC-----AGAGCGAGGAACAAGAACCGTTCACTGCTTAAGAG 373

OY 83 SerArgArgLySLysAlaenAspArgIUAryasnArgMethISapLEuanSerAla 102

Db 374 AACCGCGCGCTTAAGGCAATGACCGCGAGGAAAGCAAGATGCACAGCTTGAAATGACGC 433

OY 103 LeuSpAlaleuArgGIyValleuProThrPheProAspAlalySleuthRyle 122

Db 434 TTGGAGACCTCCCGACCGCTTCTACCGCGCTCCCGCATGACAGAACTCACAAAGTC 493

OY 123 GluThrLeuArgPheAlahIsasEnTyrlleTPralaleuthrglnthrLeuArgIlea 142

Db 494 GAACCTCGCGCTTCGCTCAATTAATCATCTGGGCACTCTCCGAGACCATATGCGATGCG 553

OY 143 Asp 143

Db 554 GAT 556

RESULT 11

LOCUS BG808248 600 bp mRNA linear EST 20-DEC-2001

DEFINITION 2083-52 Mouse B14.5 retina lambda ZAP II library Mus musculus cDNA,

ACCESSION BG808248

VERSION BG808248

KEYWORDS BG808248.1 GI:17955225

SOURCE EST.

ORGANISM Mus musculus (house mouse)

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murineae; Mus.

AUTHORS Mu,X., Zhao,S., Pershad,R., Hsieh,T.-F., Scarpa,A., Wang,S.W.,
White,R.A., Beremand,P.D., Thomas,T.L., Gan,L. and Klein,W.H.
Gene expression in the developing mouse retina by EST sequencing
and microarray analysis

TITLE Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL MEDLINE 21671825

PIRBASE 11812828

COMMENT Contact: Klein WH
Department of Biochemistry and Molecular Biology
University of Texas M.D. Anderson Cancer Center
Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
Tel: 713 792 3646
Fax: 713 790 0329

FEATURES Location/Qualifiers

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1001)
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished
COMMENT On Feb 15, 2001 this sequence version replaced gi:12869886.
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: sequef@genoscope.cns.fr, Web : www.genoscope.cns.fr
was not normalized. Library was constructed by Life Technologies, a
division of Invitrogen. This sequence belongs to sequence cluster
2626.r, Contact : Feng Liang Email : fliang@lifestech.com URL :
http://fulllength.invitrogen.com/InvitrogenCorporation 1600
Paraday Avenue Genoscope sequence ID : CS0DF035CB10QPI.
Location/Qualifiers
1..1001
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="CS0DF035YD19"
/tissue_type="FETAL BRAIN"
/dev_stage="fetal"
/clone_lib="Homo sapiens FETAL BRAIN"
/note="Organ: Brain; Vector: pCMVSPORT 6, 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-strand cDNA was digested with Not I and
cloned into the Not I and EcoRV sites of the pCMVSPORT 6
vector. Library was not normalized."
BASE COUNT 190 a 322 c 342 g 121 t 26 others
ORIGIN

Alignment Scores:
Pred. No.: 4.9e-11 Length: 1001
Score: 298.50 Matches: 81
Percent Similarity: 48.50% Conservative: 16
Best Local Similarity: 40.50% Mismatches: 40
Query Match: 26.49% Indels: 64
DB: 9 Gaps: 8

US-09-595-947E-10 (1-214) x AL540071 (1-1001)

QY 31 CysProthrSerAlaProProSerProthrArgThrProGlyAsnGys----- 46
DB 421 TGTCTCTCGATCGCGCTCTCCCGCTTGGCGGCGCTGACCCCGCTGATCAGCGCG 480
QY 47 -----AlaGluAlaGluGluGlyGlyCysArgGlyAlaProArglys 60
DB 401 ACGAAGAAAGAGAGAGAGCGCGGCGCTGACGCGGCGCGCTGCGAGCGCGGCG 540
QY 61 LeuArgAlaArgArgGlyGlyArgSer----- 69
DB 541 TTA---NGCCGGCGCAANGCGGCGCGGCGCTTCTCGCGGCGTCCGAGGCGCTCCGG 597
QY 70 -----ArgProlysSerGluLeuAla 76
DB 598 CCCGACGCGCTGCTGGTCTGTGACATGCAACGCGCGCTTCCCGGCGCGCGGCGC 657
QY 77 LeuSerLysGln-----ArgArgSerArgArglys 86
DB 658 GCTCTCCGAGGCGCCAGACGCGCCGACGCTGACGCGCATCAAGAAAGCCGTTGACTG 717
QY 87 LysAlaAsnAspArgGluArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAlaLeu 106
DB 718 AAGGCAACAACCGGAGGAGAAACCGCAACACCTCAACGCGGCACTGACCGCGCTG 777
QY 107 ArgGlyValLeuProThrPheProAspAlaLysLeuThrLysIleGluThrLeuArg 126
DB 778 CCGGAGTGTCTCCCACTTCCCGAGGAGCGCAAGCTCAACAGATCGAGACMCTGCGC 837

QY 127 PheAlaHisAsnTyrlleTrrAlaLeuThrGlnThrLeuArgIleAlaAspHisSerLeu 146
DB 838 TTCCGCCCACTAATCTGGGAGACTCAGAGACCTTCGCTCGCGGRT----- 888
QY 147 TyraLeuGluProProAlaProHisCysGlyGluLeuGlySerProGlyGlyProPro 166
DB 889 -----MAYTCCGGGCGG-GCGGCGCTCG-----GCGGCGANCGCG 920
QY 167 GlyAspTrpGlySerLeuTySerProValSerGlnAlaGlySerLeuSerProAlaAla 186
DB 921 GGG-----GGGCTCTCTTCT-----GAGCGMRTGTTKXTAGMCGCGAGCG 962

RESULT 14
LOCUS BE780690 1039 bp mRNA linear EST 20-OCT-2000
DEFINITION 601469349P1 NIH_MGC_67 Homo sapiens cDNA clone IMAGE:3872172 5',
mRNA sequence.
ACCESSION BE780690.1 GI:10201888
VERSION BE780690
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1039)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cga@bbs-remail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/ILMN at:
http://image.llnl.gov
Plate: LHM9626 row: b column: 13
High quality sequence stop: 692.
Location/Qualifiers
1..1039
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:3872172"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH MGC 67"
/note="Organ: eye; Vector: pCMV-SPORT6; Site 1: NotI;
Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.75 kb. Library constructed by Life
Technologies."
BASE COUNT 228 a 351 c 337 g 122 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 6.96e-11 Length: 1039
Score: 296.50 Matches: 95
Percent Similarity: 49.33% Conservative: 16
Best Local Similarity: 42.22% Mismatches: 66
Query Match: 26.31% Indels: 53
DB: 10 Gaps: 6

US-09-595-947E-10 (1-214) x BE780690 (1-1039)

QY 4 GlnProSer-----GlyAlaProThrValGlnValThrArgGluThrGluArgSer 20
DB 244 CAGCCCGCTTGAACCTCGACATCTCCGACTCGACTGCGCAGCAGCAGCGGCACTGACC 303
QY 21 PheProArgAlaSerGluAspGluValThrCys-----ProThr 33
DB 304 TATCCGCTTCTCTCAACC-GACGAGGAGACATGTGCGACATCTCAACAGCAGCGCTCCGCT 362

Oy		34	SerialProPsoSerProThArqThProGlyasnCysAlaGluaGlUglYlYl	53
Db		363	TCCGGCGGCCCGCCGC	386
Oy		54	CysArgIylAlaProArgIylAleuArgIla	63
Db		387	--AGGGGCGGCCCAATATCTCCCGGGCGTCTAGAGTTCCAGGGGACAGAGCAGCAG	443
Oy		64	-----ArgArgIylGIlyArgSerArgProlysSerGIleuAlaleuSerIlys	79
Db		444	CAGAGAGCGCGCGCGCGCGCGCGAGCGGGGTTCGCTCGAAGGCGTGCTCACTCG	503
Oy		80	GlnArgArgSerArgArgIylAleuAlaAsnAspArgGluArgAsnArgMetHisAspLeu	99
Db		504	CTGCGCAGAGACC CGCGCGCTCAAGGCCACCATTCGACGCCAACCGATGCACAAC TTG	563
Oy		100	A snSerAlaleuAspAlaleuArgIylValleuProThrPheProAspAspAlaleu	119
Db		564	AACGGCGCCCTGGACGCACTCGCACGCGTGCCTCG-TTCCCCGACGAACCAAGCTC	622
Oy		120	ThrlyslIegIuThrlleuArgPheAlahlsasntYrIleTrpalaleuThInThrlleu	139
Db		623	ACCAAAATCGAGACGCTGC--TTGGGCTACAA-PACACTGGGGCTCGGCCGAGACACTG	680
Oy		140	ArgIleAlaAspHisSerIleuTyralaleuGluProProAlaProHisCyEgIylGlu	159
Db		681	CGCTCGGGGATCAAAGGCTGCGCGAGAGGTGCCCGAAGCGGCTCTCGCGGMAA--	738
Oy		160	GlySerProGIylGIyProPro-----GlyAsPTrGIylSerIleu	172
Db		739	GGCGACTCGCTGGCCGACCCCAAGCGCGCAAGAAGGAGAGACTGGGGCAAGCA	798
Oy		173	TyrSerProValser-----GlnIlaGIylSerIleuSerProIlaAlaSerIleu	188
Db		799	CGCGGGCCCCCGACACTGAAGCCACAGATCCGGGCTCCGAAGAAAATAAGAGCCGGGCAC	858
Oy		189	GluGIuArgProGIyl	193
Db		859	CGACACAGCACGAGC	873
RESULT 15				
BUE12495				
LOCUS				
DEFINITION				
VERSION				
KEYWORDS				
ACCESSION				
SOURCE				
ORGANISM				
REFERENCE				
AUTHORS				
TITLE				
JOURNAL				
COMMENT				

```

BASE COUNT      131 a      244 c      264 g      91 t
ORIGIN
/db xref="taxon:10090"
/clone="UI-M-FRO-cbc-k-21-0-UI"
/tissue_type="whole brain"
/dev_stage="embryo.13.5,14.5,16.5,17.5dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH_BMAP_FRO"
/note="Organ: Brain; Vector: pYX-Asc; Site.1: EcoR I;
Site.2: Not I; The library was constructed according
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured RNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with oligo-dT
primer containing a Not I site. Double strand cDNA was
size selected according to mRNA size fraction, ligated
with EcoR I adaptor, digested with NotI and then cloned
directionally into pYX-Asc vector. The library tag
sequence located between the Not I site and the polyA tail
is AGCGAGACAG. This library was created for the University
Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
Developing Mouse Nervous System', supported by National
Institute of Mental Health (NIMH), Hemlin Chin, Ph.D.,
program coordinator."

```

BASE COUNT	131 a	244 c	264 g	91 t
ORIGIN				
Alignment Scores:				
Pred. No.:	6.28e-11			Length: 730
Score:	295.00			Matches: 78
Percent Similarity:	55.06%			Conservative: 20
Best local Similarity:	43.82%			Mismatches: 28
Query Match:	26.18%			Indels: 52
DB:	13			Gaps: 8

Oy	16	GlutThrGluArgSerPheProArgAla-----SerGluAspGluVal	29
Db	198	GAGCGCGGTAGAGATGTTCTGCAAAATCTGAGACTGTGAAGAGAAAGAGAGSTA	257
Oy	30	-----ThrCysProThrSerAlaProSerPro-----	39
Db	258	CTGATGCTGCTGGAGCTGGCTTCCCGGCTCGGACCCCTGACCTGATCTCTCCAGC	317
Oy	40	-----ThrArgThrProGlyAsnCys-----	46
Db	318	GCGGACGAGACGAGACGAGAGACTGGCGCGCGGCTCGGCGCGGCGACGTTGA	377
Oy	47	AlaGluAlaGluGlu-----GlyGlyCysArgGly	56
Db	378	GGCGAAGCGCGGCGGCGGTGACGAGCAGTCCGCGTGGGTGCGGGGGTTGGCG--	434
Oy	57	AlaProArgIysLeuArgAla-----ArgArgGlyGlyArgSerArg	70
Db	435	---CCAGGCGCGCTGCTGGGCGCTGATGACGAGTGCAGAGCGTCCGCGCTCACGG	491
Oy	71	-----ProIysSerGluLeuAlaLeuSerIysGlnArgArgSerArgArg	85
Db	492	GCGCTTCCCGAGGTGCCAGACCGCGGAGACGCTGACGCCATCAGAAAGACCCGACGG	551
Oy	86	LysIysAlaAsnAspArgIuArgAsnArgMetHisAspLeuAsnSerAlaLeuAspAla	105
Db	552	CTCAAGCGCAACACCGGAGCGCACCGCATGCAACAACCTAAACGCGCGCTGGACGCG	611
Oy	106	LeuArgGlyValLeuProThrPheProAspAspAlaLysLeuThrIysIleGluThrIleu	125
Db	612	CTGGCGCGAGTGTGCCACCTTCCCGAGAGATCCAAAGCTCACAGAGATGAGACGCTG	671
Oy	126	ArgPheAlaHisAsnIyrIleThrAlaLeuThrGlnThrIleuArgIleAlaAsp	143
Db	672	CGCTTGGCGCCACATTAATCTGGACGGCTCACCGAGACTCTGCGCTTGGCGGAC	725

Search completed: February 2, 2004, 21:35:43
Job time : 2054 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 26, 2004, 18:43:11 ; Search time 5442 Seconds

(without alignments)
10975.387 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1 ggcagtagcagagagagcag.....agagtgacctaccagtggt 1460

Sequence: IDENTITY NUC
Gapex 10-0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rnd:*
36: em_hcg_mam:*
37: em_hcg_vit:*
38: em_sy:*
39: em_hngo_hum:*
40: em_hngo_mus:*
41: em_hngo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1458.4	99.9	1491	6 A91167	A91167 Sequence 1
2	1458.4	99.9	1491	6 BD023626	BD023626 Polypept
3	1458.4	99.9	1491	10 RNRELAXT	Y10619 R. norvegicus
4	1407	96.4	258815	2 AC127817	AC127817 Rattus no
5	932.2	68.0	215050	2 AC127417	AC127417 Mus muscu
6	976.6	66.9	138070	2 AC109783	AC109783 Mus muscu
7	900.6	61.7	1861	10 AF364300	AF364300 Mus muscu
8	819.8	56.2	5567	10 MMAPATH4B	Y09167 M. musculus
9	682.4	46.7	861	6 AX698801	AX698801 Sequence
10	682.4	46.7	861	10 MMU76208	U76208 Mus musculu
11	517.4	35.4	5340	9 AF234829	AF234829 Homo sapi
12	515.8	35.3	165110	9 AL450311	AL450311 Human DNA
13	512.6	35.1	173341	2 AC021954	AC021954 Homo sapi
14	456.8	31.3	1330	9 HSA133776	A133776 Homo sapi
15	210.6	14.4	170896	2 AC011010	AC011010 Homo sapi
16	147.6	10.1	1268	6 AR023709	AR023709 Sequence
17	147.6	10.1	1268	6 AR225842	AR225842 Sequence
18	147.6	10.1	1268	6 HSU63842	U63842 Human neuro
19	147.6	10.1	1675	9 BC008687	BC008687 Homo sapi
20	147.6	10.1	1717	9 BC028226	BC028226 Homo sapi
21	147.6	10.1	79970	2 AC011428	AC011428 Homo sapi
22	147.6	10.1	134506	2 AC005738	AC005738 Homo sapi
23	147.6	10.1	151044	2 AC011432	AC011432 Homo sapi
24	145.2	9.9	1527	6 AR308547	AR308547 Sequence
25	145.2	9.9	258118	10 RNUE7777	U67777 Rattus norv
26	145.2	9.9	258118	2 AC112007	AC112007 Rattus no
27	145.2	9.9	258880	2 AC095430	AC095430 Rattus no
28	140.4	9.6	735	10 MMU67776	U67776 Mus musculu
29	140.4	9.6	738	6 AR308548	AR308548 Sequence
30	140.4	9.6	932	10 MMAPATH4C	Y09166 M. musculus
31	140.4	9.6	1315	10 MMU63841	U63841 Mus musculu
32	140.4	9.6	1333	6 AR023715	AR023715 Sequence
33	140.4	9.6	1333	6 AR225848	AR225848 Sequence
34	140.4	9.6	71538	2 AC118243	AC118243 Mus muscu
35	140.4	9.6	149268	2 AC124395	AC124395 Mus muscu
36	139.8	9.6	1385	10 MMU76207	U76207 Mus musculu
37	139.8	9.6	1412	37 MMAPATH4A	Y07621 M. musculus
38	139.8	9.6	1719	5 AB065284	AB065284 Cynops py
39	139.8	9.6	10393	10 AF303001	AF303001 Mus muscu
40	139.8	9.6	123855	2 AC102600	AC102600 Mus muscu
41	139.2	9.5	770	5 AF123884	AF123884 Gallus ga
42	139.2	9.5	790	5 GGA012659	GA012659 Gallus ga
43	139.2	9.5	1880	5 AF303000	AF303000 Gallus ga
44	139.2	9.5	2370	9 BC036847	BC036847 Homo sapi
45	139.2	9.5	6123	9 AF303002	AF303002 Homo sapi

ALIGNMENTS

RESULT 1
A91167
LOCUS A91167 1491 bp DNA linear PAT 22-JAN-2000
DEFINITION Sequence 1 from Patent WO9827206.
ACCESSION A91167
VERSION A91167.1 GI:6740202
KEYWORDS
ORGANISM Rattus sp.
Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE 1 (bases 1 to 1491)
AUTHORS Icard-Liepkalns, C., Mallet, J. and Corresponding, N.A.
JOURNAL Patent: WO 9827206-A 1 25-JUN-1998;

ICARD LIEPKAINS CHRISTINE (FR) ; MALLET JACQUES (FR)
 FEATURES
 source
 1..1491
 /organism="Rattus sp."
 /mol_type="genomic DNA"
 /db_xref="taxon:10118"
 459..1103
 /note="unnamed protein product"
 /codon_start=1
 /protein_id="CAB69414.1"
 /db_xref="GI:6740203"
 /translation="MAPPLDAPITQVQSTQPPASDHEVLSNSNTPSPITLVR
 DSEBAGDCRGTSKRLARRGRNRPSEALSRSKRSKRNKANDRENNMINLSA
 LDALRGVLPPTDAAKLTKEITLRFANHYIWAITQTLIADHSFYGPPEPVPCGELGS
 PGGSSGWMGSIYSVQAGSLSPFASIEEPFGQLVPSRPSCLLPGLTVFSDFL"
 CDS
 BASE COUNT 307 a 487 c 413 g 284 t
 ORIGIN
 Query Match 99.9%; Score 1458.4; DB 6; Length 1491;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

781 GTGTCTGCCACCTTCCCGATGACGCCAACTTAAGATGAGACCTGCGCTTCG 840
 781 GTGTCTGCCACCTTCCCGATGACGCCAACTTAAGATGAGACCTGCGCTTCG 840
 841 CCCACAACTAATTTGGGACATGACTCAGAGCTGGCCATPAGCGGACCAAGCTTTACG 900
 841 CCCACAACTAATTTGGGACATGACTCAGAGCTGGCCATPAGCGGACCAAGCTTTACG 900
 901 GCCCGAGACCCCTGTGCGCTGTGGGAGCTGGGAGACCCCGGAGGGGCTCCAGCGCG 960
 901 GCCCGAGACCCCTGTGCGCTGTGGGAGCTGGGAGACCCCGGAGGGGCTCCAGCGCG 960
 961 ACTGGGGCTATATCTACTCTCCCAAGTTTCCAAAGCTGTAGCTGAGCCCAACCTTAT 1020
 961 ACTGGGGCTATATCTACTCTCCCAAGTTTCCAAAGCTGTAGCTGAGCCCAACCTTAT 1020
 1021 TGGAGAGTTCCCTGCGCTGAGAGTGGCCAGCTCCCATCTGTCTGCTCCCGGACACC 1080
 1021 TGGAGAGTTCCCTGCGCTGAGAGTGGCCAGCTCCCATCTGTCTGCTCCCGGACACC 1080
 1081 TGGTGTCTCAGACTCTGTGGAAGGCGCCAAAGAGCCCTGGGCGGTGGCGCTGGCAG 1140
 1081 TGGTGTCTCAGACTCTGTGGAAGGCGCCAAAGAGCCCTGGGCGGTGGCGCTGGCAG 1140
 1141 AAAGGAGAGGAGTGAAGCTGTGAAATGGAAGTGTAGAGGACCTGAGCAATCTTGC 1200
 1141 AAAGGAGAGGAGTGAAGCTGTGAAATGGAAGTGTAGAGGACCTGAGCAATCTTGC 1200
 1201 CCTTCTGCGCTTCAATTAATGTAAGTCCCTGATTTTACAGAGATTTGCAAGTTCTTGC 1260
 1201 CCTTCTGCGCTTCAATTAATGTAAGTCCCTGATTTTACAGAGATTTGCAAGTTCTTGC 1260
 1261 GCTGTGCTGACCAAGGACATGTGAGCTATCTCTTAACCTTCTCAAGTGTGCGC 1320
 1261 GCTGTGCTGACCAAGGACATGTGAGCTATCTCTTAACCTTCTCAAGTGTGCGC 1320
 1321 AACTCAAACTCCCGCTCCAGAGAGAGAGACCGGTAGACTTAATAGTTGGAGACTTCC 1380
 1321 AACTCAAACTCCCGCTCCAGAGAGAGAGACCGGTAGACTTAATAGTTGGAGACTTCC 1380
 1381 ATACTTCTGTGACTCCGCGCTCTTTAAATTCGCGGCTCCAACTCAGCTTTCTCC 1440
 1381 ATACTTCTGTGACTCCGCGCTCTTTAAATTCGCGGCTCCAACTCAGCTTTCTCC 1440
 1441 AGAGTGAACCTAATCCAGTGT 1460
 1441 AGAGTGAACCTAATCCAGTGT 1460

RESULT 2
 BD023626 1491 bp DNA linear PAT 27-AUG-2002
 LOCUS
 DEFINITION
 Polypeptide belonging to the family of basic helix-loop-helix
 (bHLH) family and nucleic acid sequence corresponding thereto.
 ACCESSION
 BD023626
 VERSION
 BD023626.1 GI:22564849
 KEYWORDS
 JP 2001510464-A/1.
 SOURCE
 Rattus sp.
 ORGANISM
 Rattus sp.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 1491)
 Liepka, C.I., Mallet, J. and Ravassard, P.
 Polypeptide belonging to the family of basic helix-loop-helix
 (bHLH) family and nucleic acid sequence corresponding thereto
 Patent: JP 2001510464-A 1 31-JUL-2001;
 JOURNAL
 RHOSE POULENC ROBER SA
 OS
 Rattus sp. (rat)
 PN
 JP 2001510464-A/1
 PD
 31-JUL-2001
 PF
 19-DEC-1997 JP 1998527415


```

PR      19-DEC-1996  FR      96/115651
PI      CHRISTINE ICARD LIEPKINS,JACQUES MALLET,PHILIPPE RAVASSARD .PC
C07K4/47,A61K31/711,A61K35/76,A61K38/00,A61K48/00,A61P25/00, PC
C12N15/09,
PC      C12N15/00,A61K37/02
CC      Strandedness: Single;
CC      Topology: linear;
FH      Key      Location/Qualifiers.
          Location/Qualifiers

```

BASE COUNT	307 a	487 c	413 g	284 t
ORIGIN				

Query Match	99.9%	Score 1458.4	DB 6	Length 1491
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1459, Conservative	0	Mismatches 1	Indels 0	Gaps 0

OY	1	GCAGGTACGAGAGGAGCGAGTCCCTGAGGCCCCCGTTGCTGATTGGACCGTGGCACAAGGCA	60
Db	1	GCAGGTAGCGAAGGAGGACGTCCCTGAGGCCCCCGTTGCTGATTGGACCGTGGCACAAGGCA	60
OY	61	GCAGCCCGGACAGGCAACGCTCTGATCCGGGACAGACAGATAAAGCGTGCACAGGGACACA	120
Db	61	GCAGCCCGGACAGGCAACGCTCTGATCCGGGACAGACAGATAAAGCGTGCACAGGGGACACA	120
OY	121	CGATTAGCAGCTCAGAAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCGAGGAGCCCT	180
Db	121	CGATTAGCAGCTCAGAAAGTCCCTCTGGGTCTCACCACTGCACAGAGGCGAGGAGCCCT	180
OY	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCCAGGCGAGGAGCGCTCAGCTCAG	240
Db	181	CCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCCAGGCGAGGAGCGCTCAGCTCAG	240
OY	241	CAAACTTCGAAGCGACGAGAGGGGTTCAGCTATCCACGCTGCTTGACTCTGACCAACC	300
Db	241	CAAACTTCGAAGCGACGAGAGGGGTTCAGCTATCCACGCTGCTTGACTCTGACCAACC	300
OY	301	GCAGCTCTGTCTTTTGGCCGCGAAGTAACTTAGGTAACTTTAGGAACCTCCAAAGG	360
Db	301	GCAGCTCTGTCTTTTGGCCGCGAAGTAACTTAGGTAACTTTAGGAACCTCCAAAGG	360
OY	361	TAGAAGAGGGGAGTGGGTGGCGTACTCTAATGCCGCTGAGTAGACTCTAAGTCAGG	420
Db	361	TAGAAGAGGGGAGTGGGTGGCGTACTCTAATGCCGCTGAGTAGACTCTAAGTCAGG	420
OY	421	ACTGTACACACCCCTTCATTTTTCACCACTGAGATGGCGCTCATCCCTTGGATG	480
Db	421	ACTGTACACACCCCTTCATTTTTCACCACTGAGATGGCGCTCATCCCTTGGATG	480
OY	481	CGCCCAACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGAGCTCTGACACACG	540
Db	481	CGCCCAACCATCCAGTGTCCCAAGAGACCAGCAACCTTTCCCGAGCTCTGACACACG	540
OY	541	AAGTGTCAAGTTCGAATTCACACCCACCTTAGGCCCACTCTGTAACGAGGACCTGTCG	600
Db	541	AAGTGTCAAGTTCGAATTCACACCCACCTTAGGCCCACTCTGTAACGAGGACCTGTCG	600
OY	601	AAGCAGAAGCAGGTGACTGCGAGGGGACATCGAGGAAGTCCGTTGCGGCGCGAGGCG	660
Db	601	AAGCAGAAGCAGGTGACTGCGAGGGGACATCGAGGAAGTCCGTTGCGGCGCGAGGCG	660
OY	661	GCAACAGGCGCCAAAGACGAGTTGGCACTAGACGAAGCAGGACGGAAGCCGCGCAAGAGG	720
Db	661	GCAACAGGCGCCAAAGACGAGTTGGCACTAGACGAAGCAGGACGGAAGCCGCGCAAGAGG	720
OY	721	CCAAAGACCGGGAGCGCAACCGCATGCAACACTTAACCTCGCGCTGATGCGCTGCGG	780
Db	721	CCAAAGACCGGGAGCGCAACCGCATGCAACACTTAACCTCGCGCTGATGCGCTGCGG	780
OY	781	GTTGCTCTGCCACTTCCCGGATGACGCCAAACTTACAAAGATCGAGACCTTGCCCTTG	840

Db	781	GTGTCCTGCCCCACTTCCCGGATAGACGCAAACTTACAAAGATGAGACCCTGGCGTTCCG	840
QY	841	CCCAACAATCAATTTGGGCACTGACTGACAGACGCTGGCGATATGCGGACCAACAGCTTTCACG	900
Db	841	CCCAACAATCAATTTGGGCACTGACTGACAGACGCTGGCGATATGCGGACCAACAGCTTTCACG	900
QY	901	GGCCCCGAGCCCCCTGTCGCTGCGTGGGGAGCGGGGAAGCCCGGGAGGGGGGCTCCAGCGGCG	960
Db	901	GGCCCCGAGCCCCCTGTCGCTGCGTGGGGAGCGGGGAAGCCCGGGAGGGGGGCTCCAGCGGCG	960
QY	961	ACTGGGGGCTCATCTACTCCCAAGTTTCCCAAGCTGTGAGCTGAGCCCAACAGCTTCAT	1020
Db	961	ACTGGGGGCTCATCTACTCCCAAGTTTCCCAAGCTGTGAGCTGAGCCCAACAGCTTCAT	1020
QY	1021	TGAGAGAGTTCCCTGGCGCTGCAAGGTGCCAGCTCCCATCCTGTCTGTCCTCCCGGGACACC	1080
Db	1021	TGAGAGAGTTCCCTGGCGCTGCAAGGTGCCAGCTCCCATCCTGTCTGTCCTCCCGGGACACC	1080
QY	1081	TGGAGTCTCAGACTTCCTGTGAAAGGGGCCCAACAGGCGCTGGGCGGGGCGCTGGCAG	1140
Db	1081	TGGAGTCTCAGACTTCCTGTGAAAGGGGCCCAACAGGCGCTGGGCGGGGCGCTGGCAG	1140
QY	1141	AAAGGGAGGGAGTCAAGAGCTGCTCGAATATGAAAGTATGTGAGGCACTCGAGACTTCGCG	1200
Db	1141	AAAGGGAGGGAGTCAAGAGCTGCTCGAATATGAAAGTATGTGAGGCACTCGAGACTTCGCG	1200
QY	1201	CCCTTCTGCGCTTCAATTAATGTCAGGTCCCTGATTTTAACCAAGATTCGCAAGTTCCTTGCT	1260
Db	1201	CCCTTCTGCGCTTCAATTAATGTCAGGTCCCTGATTTTAACCAAGATTCGCAAGTTCCTTGCT	1260
QY	1261	GCTGTGCGTGGACAAGAGATTTTCAGGCGGTATCTCCCTTAACCGCTCCAGTGTGGCC	1320
Db	1261	GCTGTGCGTGGACAAGAGATTTTCAGGCGGTATCTCCCTTAACCGCTCCAGTGTGGCC	1320
QY	1321	ACCTCAAACTCCCGCTCCAGACAGAGAGACCGTAGACTTAATATGTGTGGAGACTCC	1380
Db	1321	ACCTCAAACTCCCGCTCCAGACAGAGAGACCGTAGACTTAATATGTGTGGAGACTCC	1380
QY	1381	ATACTTCTGTGTACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAAACCAACGCTTTCCTCC	1440
Db	1381	ATACTTCTGTGTACTCCGCCCTCTTTCAAATCTGCGGGCCTCCAAACCAACGCTTTCCTCC	1440
QY	1441	AGAGTGACCTAATCCAGTGT 1460	
Db	1441	AGAGTGACCTAATCCAGTGT 1460	

RESULT 3			
LOCUS	RNRELAXT	1491 bp	mRNA linear
DEFINITION	R.norvegicus mRNA for transcriptional regulator, Relax.		ROD 06-MAY-1997
ACCESSION	Y10619		
VERSION	Y10619.1	GI:2072737	
KEYWORDS	Relax; transcriptional regulator.		
SOURCE	Rattus norvegicus (Norway rat)		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.		
REFERENCE	1		
AUTHORS	Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.		
TITLE	Relax, a novel rat BHLH transcriptional regulator transiently expressed in the ventricular proliferating zone of the developing central nervous system		
JOURNAL	J. Neurosci. Res. 48 (2), 146-158 (1997)		
MEDLINE	97276380		
PubMed	9130143		
REFERENCE	2 (bases 1 to 1491)		
AUTHORS	Ravassard, P.		
TITLE	Direct Submission		
JOURNAL	Submitted (20-JUN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERVL, Hopital de la Plie Salpêtrière, 83 Bd. de l'Hopital, F-75013		

FEATURES		Paris, FRANCE
source	Location/Qualifiers	
	1..1491	
	/organism="Rattus norvegicus"	
	/mol_type="mRNA"	
	/strain="Wistar"	
	/db_xref="taxon:10116"	
	/clone_id="RT-PCR"	
	459..1103	
	/function="transcriptional regulator"	
	/codon_start=1	
	/product="Relax"	
	/protein_id="CAA71630.1"	
	/db_xref="gi:2072738"	
	/db_xref="SPTRMBL:O08718"	
	/translation="MAPEHPLAPRTQVSQETQQPEPAGSDHEVLSSNTPSPFLVPERDSENEADCGRTSRKLAARRGRRPKSEIALSKORSRRKKANDRRNMHNINSLDLRAPPTPEPDAKLTKIETLRPAHYIWALETADHSFYGPPEPVCSELGSPGGSSGDMGSIVSPVQASISPASIEEPFGIVPSPSCLLPLGLVFSDFL"	
BASE COUNT	307 a 487 c 413 g 284 t	
ORIGIN		
Query Match	99.9%; Score 1458.4; DB 10; Length 1491;	
Best Local Similarity	99.9%; Pred. No. 0;	
Matches 1459; Conservative	0; Mismatches 1; Indels 0; Gaps 0;	
QY	1 GCAGGTAGCGAGAGGACGATGTCCTGGGCCCCGTTGCTGATTTGGCCCCGTGGCACAGCA	60
DB	1 GCAGGTAGCGAGAGGACGATGTCCTGGGCCCCGTTGCTGATTTGGCCCCGTGGCACAGCA	60
QY	61 GCAGCCCGGACGGACGGCTCTGTGTCGGGGACAGACAGTAAAGGTGCGAGGGGACACA	120
DB	61 GCAGCCCGGACGGACGGCTCTGTGTCGGGGACAGACAGTAAAGGTGCGAGGGGACACA	120
QY	121 CGATTAGACGCTCAGAAAGTCCCTCTGGGCTCTCACCACTGCACAGAGCGCGAGACCCCT	180
DB	121 CGATTAGACGCTCAGAAAGTCCCTCTGGGCTCTCACCACTGCACAGAGCGCGAGACCCCT	180
QY	181 CCGAGCTTCTTTGTGCTCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACAGTCA	240
DB	181 CCGAGCTTCTTTGTGCTCTCCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACAGTCA	240
QY	241 CCAAACTTCGAAAGCGACGAGAGGGGTACGCTATCCACCGCTGTTGACTCTGACACACC	300
DB	241 CCAAACTTCGAAAGCGACGAGAGGGGTACGCTATCCACCGCTGTTGACTCTGACACACC	300
QY	301 GCAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACTTTAGAGAACTCCCAAAAGG	360
DB	301 GCAGCTCTCTGTTCTTTTGGAGCCCGGAGTAACTTTAGAGAACTCCCAAAAGG	360
QY	361 TAGAAGAGGGAGTGGGTGGGCGTACTTAACTCCGCGTGAAGTAACTTAAAGTCAAG	420
DB	361 TAGAAGAGGGAGTGGGTGGGCGTACTTAACTCCGCGTGAAGTAACTTAAAGTCAAG	420
QY	421 ACTGTCAACACCCCTTCATTTTCCCAACTCAGATGGCGCTCATCCCTTGATG	480
DB	421 ACTGTCAACACCCCTTCATTTTCCCAACTCAGATGGCGCTCATCCCTTGATG	480
QY	481 CGCCACCATCCAGTGTCCCAAGAGACCCAGAACCCCTTCCCGGAGGCTCGACACAG	540
DB	481 CGCCACCATCCAGTGTCCCAAGAGACCCAGAACCCCTTCCCGGAGGCTCGACACAG	540
QY	541 AAGTCTCAGTTCCAATTCACCCCACTTAGCCCACTTCGTACCGAGGACTCTCCG	600
DB	541 AAGTCTCAGTTCCAATTCACCCCACTTAGCCCACTTCGTACCGAGGACTCTCCG	600
QY	601 AAGCAGAGCAGGTGACTGCGAGGGACATTCAGAGGAAGCTCCGTGCGCGCGAGAGGC	660
DB	601 AAGCAGAGCAGGTGACTGCGAGGGACATTCAGAGGAAGCTCCGTGCGCGCGAGAGGC	660
QY	661 GGAACAGCGCCAAAGCGAGTGGACCTAGAGAGAGAGAGGACCGCGCCAAAGAG	720
DB	661 GGAACAGCGCCAAAGCGAGTGGACCTAGAGAGAGAGAGGACCGCGCCAAAGAG	720

OY		721	CGAACGACGGGAGCGCAACCGCATGGAACAACCTTAATCTCGCGTGTGATTCGCGTGCGC	780
Db		721	CMAACGACCGGAGACCGCAACCGCATGGAACAACCTTAATCTCGCGTGTGATTCGCGTGCGC	780
OY		781	GTCGTCGACCACCTTCCGGATGACGCACAACTTACMAAAGATGAGACCCCTGCGCTTGC	840
Db		781	GTCGTCGACCACCTTCCGGATGACGCACAACTTACMAAAGATGAGACCCCTGCGCTTGC	840
OY		841	CCCACAATCAATTATTTGGGGCACTGACTAGAAGCTGTGGCATATAGCGGACCAAGCTTCTTAGC	900
Db		841	CCCACAATCAATTATTTGGGGCACTGACTAGAAGCTGTGGCATATAGCGGACCAAGCTTCTTAGC	900
OY		901	GCCCGGAGCCCCCTGTGCTCCCTGTGAGGGAGCTGGGCAACCCGGGAGGGGGGCTCCAGCGAGC	960
Db		901	GCCCGGAGCCCCCTGTGCTCCCTGTGAGGGAGCTGGGCAACCCGGGAGGGGGGCTCCAGCGAGC	960
OY		961	ACTGGGGCTCTATCTACTCTCCCAAGTTTTCCCAAGCTGTGTAGCTTGAGCCCAACAGCTTCAT	1020
Db		961	ACTGGGGCTCTATCTACTCTCCCAAGTTTTCCCAAGCTGTGTAGCTTGAGCCCAACAGCTTCAT	1020
OY		1021	TGGAGGAGTTCCCTGGGCTGCAAGTGTCCCAAGCTCCCATCTGTCTGTCTCCCGGGACACC	1080
Db		1021	TGGAGGAGTTCCCTGGGCTGCAAGTGTCCCAAGCTCCCATCTGTCTGTCTCCCGGGACACC	1080
OY		1081	TGGAGTTCTCAGACTTCTTTGTAAGGGGCCCAACAGGCCCTGTGGCGGTGGCGCTGGCAG	1140
Db		1081	TGGAGTTCTCAGACTTCTTTGTAAGGGGCCCAACAGGCCCTGTGGCGGTGGCGCTGGCAG	1140
OY		1141	AAGGAGGAGGAGTCAAGACTGTCTGAATATGGAAGTGTGAGGCACTCGAGCATCTTCGC	1200
Db		1141	AAGGAGGAGGAGTCAAGACTGTCTGAATATGGAAGTGTGAGGCACTCGAGCATCTTCGC	1200
OY		1201	CCCTTCTGAGCTTCAATTAAGTACAGGTCCCTGTATTAACAGAGATTGCGACAGTTCCTTGC	1260
Db		1201	CCCTTCTGAGCTTCAATTAAGTACAGGTCCCTGTATTAACAGAGATTGCGACAGTTCCTTGC	1260
OY		1261	GCTGTGGCTGCACAAAGATATTTCAGAGCTGATCTCTTTAACCCCTCTTCAGTGTGGCC	1320
Db		1261	GCTGTGGCTGCACAAAGATATTTCAGAGCTGATCTCTTTAACCCCTCTTCAGTGTGGCC	1320
OY		1321	ACTCTAAATCTCCGCTCTCCAMGACAGAGAGAGCCGTAGACTAAATATGTTGGGAAGACTCCC	1380
Db		1321	ACTCTAAATCTCCGCTCTCCAMGACAGAGAGAGCCGTAGACTAAATATGTTGGGAAGACTCCC	1380
OY		1381	ATACTCTCTGGTGACTCGGCCCTCTTTCAATCTGCGGGGCTCCCAACACCGCTTCTTCC	1440
Db		1381	ATACTCTCTGGTGACTCGGCCCTCTTTCAATCTGCGGGGCTCCCAACACCGCTTCTTCC	1440
OY		1441	AGAGTGACCTAATCAAGTGT 1460	
Db		1441	AGAGTGACCTAATCAAGTGT 1460	
RESULT 4				
AC127817		258815 bp	DNA	linear HTG 19-NOV-2002
LOCUS				
DEFINITION	Rattus norvegicus clone CH230-259G16,	WORKING DRAFT SEQUENCE, 3		
ACCESSION	unordered pieces.			
VERSION	AC127817.3	GI:25077905		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
SOURCE	Rattus norvegicus (Norway rat)			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
	Rattus.			
REFERENCE	1 (bases 1 to 258815)			
AUTHORS	Mundy,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J., Allen,C., Allen,H., Albrooks,S., Amin,A., Angiano,D., Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H., Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F., Biswalio,K., Blair,T., Blankenburg,K., Blyth,P., Brown,M.,			

Bryant, N., Buhay, C., Butcher, K., Butrell, R., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Cesar, H., Center, A.,
 Chacko, V., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carrroll, L., De Ande, C., Dederich, D.,
 Delgado, O., Denison, S., Desmo, C., Ding, Y., Dim, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durkin, K., Duvall, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Gatta, R., Garcia, A., Garner, T., Garza, M.,
 Gbureguogis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlik, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M.,
 Hollins, B., Howells, S., Huliy, S., Hume, J., Idididd, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
 Karpach, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kwis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, M., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshewa, L., Louiseged, H., Lozano, R. T., Lu, X., Ma, J.,
 Maheswari, M., Mahindartne, M., Mamoud, M., Malloy, K., Mangum, A.,
 Mangun, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
 Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L.,
 Puzo, M., Quirroz, J., Rachlin, E., Reeves, K., Reiter, M. A., Reigh, R.,
 Rella, B., Rellay, M., Ren, Y., Reuter, R., Richards, S., Riggs, F.,
 Rivers, C., Rodkey, T., Rojce, A., Rose, M., Rose, R., Ruiz, S.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smaj, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steidle, M., Strong, R., Sutton, A., Svetek, A., Taber, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trojes, Z., Uman, K.,
 Wang, R., Vera, V., Villanana, D., Walron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wlecyk, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstein, G., and Gibbs, R. A.
 National Human Genome Research Institute
 Bethesda, Maryland 20892-0901
 United States of America
 2003

```

Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

-----
Project Information
Center project name: GZXS
Center clone name: CH230-259G16
Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228242; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
-----

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.
*
* Location/Qualifiers
* 1..258815
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-259G16"
* 1..1609
misc_feature
/note="wgs_end_extension
clone_end:5p6"
complement(2865..3659)
/note="Clone boundary
clone_end:5p6
site:_"
end_sequence:BF169379"
224098..225157
/note="Clone boundary
clone_end:T7
site:_"
end_sequence:BF169378"
BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others
ORIGIN
Query Match 96.4%; Score 1407; DB 2; Length 258815;
Best Local Similarity 99.1%; Pred. No. 2,5e-305;
Matches 1446; Conservative 0; Mismatches 10; Indels 3; Gaps 3;
4 GGTAGCGAGAGGAGAGAGTCTCGGGCCCCCGTGTGTAATTGGCCGTGGCACAGGACGCA 63
Db 247404 GGGCTCGCTGAGCGAGTCCCTCGGGCCCCCGTGTGTAATTGGCCGTGGCACAGGACGCA 247463
QY 64 GCGCGGACGACGCTCCCTGTCGCGGGAGAGACGATTAACGTCGACAGGGG-ACACAG 122
Db 247464 GCGCGGACGACGCTCCCTGTCGCGGGAGAGACGATTAACGTCGACAGGGGACACAG 247522
QY 123 ATTAGAGCTCAAGAGTCCCTCTGGGTCTCACTGACACAGAGGCGAGAGCCCTCC 182
Db 247524 ATTAGAGCTCAAGAAATCCCTTGGGTCTCACTGACACAGAGGCGAGAGCCCTCC 247588
QY 183 GAGCTTCTTGTGCTCTCCAGACGAATTTACTCAGGCGAGGCGCGCTGCAGTTCAGCA 242
Db 247584 GAGCTTCTTGTGCTCTCCAGACGAATTTACTCAGGCGAGGCGCGCTGCAGTTCAGCA 247643
QY 243 AACTTCGAAGGACGAGGGGTTCACTTATCCACCGGCTTGACTCTGACC-ACCG 301

```

```

Db 247644 AAACCTTGAAACGAGAGAGGGGTTTCAGCTATCCACCGCTGCTTGAACCAACCCG 247703
Qy 302 CAGCTCTCTGTTCTTTTGAAGCCCGAGTAATAGTAATATTGAAGACTCCAAAGAGT 361
Db 247704 CAGCTCTCTGTTCTTTTGAAGCCCGAG-MACTAGTAACTTTAGGCCCTCCAAAGAGT 247762
Qy 362 AGAAGAGGGAGTGGTGGGCGTACTAGTCCGCGGTGAGTGAACCTTAAAGTCAAGA 421
Db 247763 AGAAGAGGGAGTGGTGGGCGTACTAGTCCGCGGTGAGTGAACCTTAAAGTCAAGA 247822
Qy 422 CTGTCAACACCCCTTTCATTTTTCACAACTAGAGATGGCGCTCATCCCTTGATGC 481
Db 247823 CTGTCAACACCCCTTTCATTTTTCACAACTAGAGATGGCGCTCATCCCTTGATGC 247882
Qy 482 GCCCACCATCCAAATGTCTCCAAAGAGACCAGCAACCTTTCCGAGAGCTTGACCAAGA 541
Db 247883 GCCCACCATCCAAATGTCTCCAAAGAGACCAGCAACCTTTCCGAGAGCTTGACCAAGA 247942
Qy 542 AGTGTCTAGTTCAATTCACACCCCACTAGCCCACTCTGTACGAGGGACTGTCCGA 601
Db 247943 AGTGTCTAGTTCAATTCACACCCCACTAGCCCACTCTGTACGAGGGACTGTCCGA 248002
Qy 602 AGCAGAAAGCAGTATCTGCGAGAGCAATCGAGAGTCTCGTGGCGGCGCGAGAGGCG 661
Db 248003 AGCAGAAAGCAGTATCTGCGAGAGCAATCGAGAGTCTCGTGGCGGCGCGAGAGGCG 248062
Qy 662 CAACAGGCGCCAAAGAGCGAGTTGGCACTGAGAGCAGAGCAGAGCGGCGAGAAAGGC 721
Db 248063 CAACAGGCGCCAAAGAGCGAGTTGGCACTGAGAGCAGAGCAGAGCGGCGAGAAAGGC 248122
Qy 722 CAACAGCAGGAGAGCGCAACCGCATGCAACAACCTTAACTCCGCGTGAATGCGCGCG 781
Db 248123 CAACAGCAGGAGAGCGCAACCGCATGCAACAACCTTAACTCCGCGTGAATGCGCGCG 248182
Qy 782 TGTCTCTGCGCCCTTCCCGAGTGAAGCCCAACTTAAGATGAGAACCTTCGCGCTTGC 841
Db 248183 TGTCTCTGCGCCCTTCCCGAGTGAAGCCCAACTTAAGATGAGAACCTTCGCGCTTGC 248242
Qy 842 CCACAACTACATTTGGGCACTGAGCTCAGACGCTGCGATAGCGAGCAACAGCTTCTACG 901
Db 248243 CCACAACTACATTTGGGCACTGAGCTCAGACGCTGCGATAGCGAGCAACAGCTTCTACG 248302
Qy 902 CCCCAGAGCCCTTGTGCGCTGTGGGAGCTGGAAGCCCGGAGAGGGGCTTCAGCGGGA 961
Db 248303 CCCCAGAGCCCTTGTGCGCTGTGGGAGCTGGAAGCCCGGAGAGGGGCTTCAGCGGGA 248362
Qy 962 CTGGGGCTCTATCTACTGCCAGTTTCCCAAGCTGTGTAAGCTGAGCCCCACAAGCTTACT 1021
Db 248363 CTGGGGCTCTATCTACTGCCAGTTTCCCAAGCTGTGTAAGCTGAGCCCCACAAGCTTACT 248422
Qy 1022 GGAGGAGTTCCCTGGCCCTGAGAGTGGCCAGCTCCCATCTGTCTGCTCCGCGGACCCCT 1081
Db 248423 GGAGGAGTTCCCTGGCCCTGAGAGTGGCCAGCTCCCATCTGTCTGCTCCGCGGACCCCT 248482
Qy 1082 GGTGTCTCAGACTTCTTGTGAAGGCGCCAAACAGGCTTGGGCGTGGCGCTGAGAGA 1141
Db 248483 GGTGTCTCAGACTTCTTGTGAAGGCGCCAAACAGGCTTGGGCGTGGCGCTGAGAGA 248542
Qy 1142 AAGGAGGAGTCTCAGAGTGTCTGAATGAGAGTGTGAGGCACTTGACACTCTGCGC 1201
Db 248543 AAGGAGGAGTCTCAGAGTGTCTGAATGAGAGTGTGAGGCACTTGACACTCTGCGC 248602
Qy 1202 CCTTCTGGCTTCTTAATGTCAGGTCCCTGAATTAACAGAGATTGGCAAGTTCCTGTG 1261
Db 248603 CCTTCTGGCTTCTTAATGTCAGGTCCCTGAATTAACAGAGATTGGCAAGTTCCTGTG 248662
Qy 1262 CTGTGCGTGCAACAAGATATTGCAAGCTGATCTCTTAACCTCTCTCAGTGTGCGCA 1321
Db 248663 CTGTGCGTGCAACAAGATATTGCAAGCTGATCTCTTAACCTCTCTCAGTGTGCGCA 248722
Qy 1322 CCTCAAACTCCCGCTCCAAAGAGAGAGAGCGCTGAGCACTAAATATGTTGGAGACTCCCA 1381
Db 248723 CCTCAAACTCCCGCTCCAAAGAGAGAGAGCGCTGAGCACTAAATATGTTGGAGACTCCCA 248782

```

```

Qy 1382 TACTTCTGTGATCTCGCCCTTCTTCAATCTGCGGCGCTCCAAACCGCTTTCTCA 1441
Db 248783 TACTTCTGTGATCTCGCCCTTCTTCAATCTGCGGCGCTCCAAACCGCTTTCTCA 248842
Qy 1442 GAGTGACTTAATCCAGTGT 1460
Db 248843 GAGTGACTTAATCCAGTGT 248861

RESULT 5
AC127417
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

AC127417 215050 bp DNA linear HTG 19-OCT-2002
Mus musculus chromosome UNK clone RP23-459W2, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC127417
AC127417.2 GI:24137619
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (15-JUL-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 215050)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (19-OCT-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On Oct 19, 2002 this sequence version replaced gi:21759524.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web Site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project Information
Center project name: M.BA0459M02

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.99019
Consensus quality: 212279 bases at least Q40
Consensus quality: 213137 bases at least Q30
Consensus quality: 213590 bases at least Q20
Insert size: 192000; agarose-fp
Insert size: 217944; sum-of-ctnigs
Quality coverage: 11.89 in Q20 bases; agarose-fp
Quality coverage: 10.54 in Q20 bases; sum-of-ctnigs

-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 4227: contig of 4227 bp in length
* 4228 4327: gap of unknown length
* 4328 16588: contig of 12262 bp in length
* 16590 16689: gap of unknown length
* 16690 27318: contig of 10629 bp in length

```

* 27319 27418: gap of unknown length.
* 27419 49797: contig of 22379 bp in length
* 49798 49897: gap of unknown length
* 49898 79378: contig of 29481 bp in length
* 79379 79478: gap of unknown length
* 79479 132595: contig of 53117 bp in length
* 132596 132695: gap of unknown length
* 132696 215050: contig of 82355 bp in length.
Location/Qualifiers
1. .215050
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP23-459M2"
1. .4227
misc_feature
/note="assembly_name:Contig31"
4328.16589
misc_feature
/note="assembly_name:Contig32"
16690.27318
misc_feature
/note="assembly_name:Contig33"
27419.49797
misc_feature
/note="assembly_name:Contig34"
49898.79378
misc_feature
/note="assembly_name:Contig35"
79479.132595
misc_feature
/note="assembly_name:Contig36"
132696.215050
misc_feature
/note="assembly_name:Contig37"
132696.215050
BASE COUNT 56024 a 52513 c 52885 g 53012 t 616 others
ORIGIN

Query Match 68.0% Score 992.2; DB 2; Length 215050;
Best Local Similarity 83.9% Pred. No. 2.8e-212;
Matches 1275; Conservative 0; Mismatches 178; Indels 66; Gaps 11;
QY 4 GGTAGCAGAGAGACAGTCCCTGGGCCCCCGTGTCTGATTGGCCCGTGGCAGAGCAGCA 63
Db 205068 GGGCTCGGTGGCCAGCCCTTG66CCCCCTCGGTATTGGCCCGTGGTCAAGCAGCA 205127
QY 64 GCCCGGAGGAGCGCTCTGTGTC-CGGGAGAGCAGATTAAGCGTGGAGGGAGACACAG 122
Db 205128 GCCCGGAGGAGCGCTCTGTGTCGGGGGAGAGAGCAGATTAAGCGTGGAGGGAGACACAG 205187
QY 123 ATT-----AGCAGCTCAGAGTCCCTGGGCTCAGCAGCTGCA-CAGAGGGCGAGAGCCC 177
Db 205188 ACTTGACATGAGCTCAGAAATCCCTGTGGGTCTATCTCAGAGAGTGTGAGTACT 205247
QY 178 CTTCCAGAGCTTCTTGTGCTCCAGACGCAATTACTCCAGGAGGAGCGCTGACACT 237
Db 205248 CTTGGAGCTTTTCTAGCACTTCAGAGCAATTACTCCAGGAGGAGCGCTGACAGTT 205307
QY 238 CAGCAAACTTGAAGCAGAGAGGGTTCACTATCCACCGTCTGTACTTGACC- 296
Db 205308 TAGCAGAACTTCAAGGAGAGAGAGGCTCAGCTATCCACTGTGCTTGACATGACC 205367
QY 297 -----ACCGCAGCTCTGTCTTTTGAAGCCGGAGTA 330
Db 205368 TATCCACTGCTCTTGTCTGACTGACTGCTGTCTCTTCTTCTTGTGAGTGGAG- 205426
QY 331 ACTAGTAACTTTAGAACTCCAAAGGTAGAGAGGGAGTG-----GGTGG 380
Db 205427 ACTAGTAACTTTAGAACTCCAAAGGTAGAGAGGGAGGGGGGTGTGTGGG 205486
QY 381 GCGTACTTACTCCCGGTGAGTGAAGCTTAACTCAGAGACTGTCAACCCC--CCTTC 438
Db 205487 GGATATCTGTGCTCCCGGTGAGTGAAGCTTAACTCAGAGAGGCTGGACACACACACCTTC 205546
QY 439 CATTTTCCCAACTCAGATGAGGCGCTCATCCCTTGGAGCGGCCCAACCAATCCAAAGT 498
Db 205547 CATTTTCCCAACTCAGATGAGGCGCTCATCCCTTGGAGCGCTCAACCAATCCAAAGT 205606
QY 499 CCCAAGAGACCAGCAACCTTTCGGGAGCTCGAGCAGAGTGTCTCAATT 558

Db 205607 CCCAGAGACAAACAACCTTTCCCGAGCTTCGACCAAGTGTCTCAATT 205666
QY 559 CCACCCACCTAGCCCACTCTGTACAGAGGAGCTGCTCCGAAGCAGAGAGTACT 618
Db 205667 CCACCCACCTAGCCCACTCTGTACAGAGGAGCTGCTCCGAAGCAGAGTACT 205726
QY 619 GCCAGAGCATGAGAGAACCTCTGTGCGGCGCGAGGAGCGCAACAGGCCAAGAGCG 678
Db 205727 GCCAGAGCATGAGAGAACCTCTGTGCGGCGCGAGGAGCGCAACAGGCCAAGAGCG 205786
QY 679 AGTTGGCACTGAGAGACAGCGCAAGAGCGGCGCAAGAGGCCCAACGACCGGAGCGCA 738
Db 205787 AGTTGGCACTGAGAGACAGCGCAAGAGCGGCGCAAGAGGCCCAACGACCGGAGCGCA 205846
QY 739 ACCGATGACAAACCTTAATCCGCGGTGATGGGCGCGGCGGCTCTGCCACTTCC 798
Db 205847 ATGCAATGACAAACCTTAATCCGCGGTGATGGGCGCGGCGGCTCTGCCACTTCC 205906
QY 799 CGATGACGCGCAACTTACAAAGATCGAGACCTGCGCTTCCGCAACATCAATTGGG 858
Db 205907 CGATGACGCGCAACTTACAAAGATCGAGACCTGCGCTTCCGCAACATCAATTGGG 205966
QY 859 CACTGACTCAGAGCTGCGCATAGCGGACCAAGCTTCTACGGCCCCGAGCCCCCTGTGC 918
Db 205967 CACTGACTCAGAGCTGCGCATAGCGGACCAAGCTTCTACGGCCCCGAGCCCCCTGTGC 206026
QY 919 CTTGTGGGAGCTGGGAGAGCCCGGAGAGGGGCTCCAGCGCGCATGAGGCTTACTACT 978
Db 206027 CTTGTGGAGCTGGGAGAGCCCGGAGAGTGGCTTCAAGCGGAGTGGGCTTACTACT 206086
QY 979 CCCAGTTTCCCAAGCTGAGCTGAGCTGAGCCCAAGCGCTTATGAGAGATTCCTGGCC 1038
Db 206087 CCCAGTTTCCCAAGCGGATTAAGCTGAGCCCAAGCGCTTATGAGAGATTCCTGGCC 206146
QY 1039 TGAAGTGCAGCTCCCATCTGTGTCTCCGAGCACCTGTGTCTCAGACTTCT 1098
Db 206147 TGAAGTGCAGCTCCCATCTGTGTCTCCGAGCACCTGTGTCTCAGACTTCT 206206
QY 1099 TGTGAAGAGCTGT 206266
Db 206207 TGTGAAGAGCTGT 206326
QY 1159 CTGTCTGAATGGAAGTATGAGAGCACTGAGACATCTGCCCTTCTGCTTCAATTA 1218
Db 206267 CTGTCTGAATGGAAGTATGAGAGCACTGAGACATCTGCCCTTCTGCTTCAATTA 206326
QY 1219 GTCAAGTCCCT-----GATTTAACAGAAATGCAAGTTCTTGTCTGTGT 1265
Db 206327 GTCAAGTCCCTGAGCCCTTCACAGGGCTTAAGGTTCTCATGTGTATC--CCTTGTGT 206384
QY 1266 GCGTGCACAAAGATATGAGAGCTGATCTCTTAAACCTCTCTCAGTGTGACACTC 1325
Db 206385 CTGGCAGACAGCAATTTGGGGCTGTCTTCTTAACTCTCTGTGTGAGCAGCAATC 206444
QY 1326 AAATCCCGCTCCAGAGAGAGAGCGCTAGCACTAAATAGTTGGAGACTCCATACT 1385
Db 206445 AAATCCCGCTCCAGAGATTTGAGATGATGAGCACTACTAGTT--GAGACTCCATACT 206503
QY 1386 TCTGTGATCTCGGCTCTTTCAAATCTGCGGCT-----CCAAACACGCTTCTCCA 1441
Db 206504 TCTGTGATCTCGGCTCTTTCAAATCTGCGGCTCCGAGCATCATCACTTTTCCA 206563
QY 1442 GAGTGAACCTAATCAAGT 1460
Db 206564 GAGTGAACCTAATCAAGT 206582

RESULT 6
AC109783/c 138070 bp DNA linear HTG 07-FEB-2002
LOCUS AC109783
DEFINITION Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
unordered pieces.

ACCESSION	ACT09783
VERSION	ACT09783.1 GI:18581594
KEYWORDS	Htg; HTGS_PHASE1; HTGS_DRAFT.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	Zuvareh,T., McCombie,W.R., Baker,J.P., Ballja,V., Dedhia,N.N., de la Baetide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A., Miller,B., Miller,S., Nascimento,L.U., O'Shaughnessy,A.L., Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and Zuvareh,T..
TITLE	Mouse Genomic Sequence
JOURNAL	Unpublished
AUTHORS	2 (bases 1 to 138070)
TITLE	McCombie,W.R.
JOURNAL	Direct Submission
COMMENT	Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, NY 11724, USA ----- Genome Center ----- Laboratory Center code: CSHL. Web site: http://www.cshl.org/genseq Contact: mcombie@cshl.org ----- Project Information ----- Project name: RP23-121F10 Clone name: RP23-121F10 Insert size: 173000; agarose-fp Insert size: 141616; sum-of-contigs Quality coverage: 4.00 in Q20 bases; agarose-fp Quality coverage: 3.70 in Q20 bases; sum-of-contigs ----- * NOTE: This is a 'working draft' sequence. It currently consists of 17 contigs. The true order of the pieces * is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved. ----- * 1 17785: contig of 17785 bp in length * 17786 17874: gap of unknown length * 17875 30356: contig of 12482 bp in length * 30357 30444: gap of unknown length * 30445 42306: contig of 11862 bp in length * 42307 42394: gap of unknown length * 42395 53598: contig of 11204 bp in length * 53599 53686: gap of unknown length * 53687 64435: contig of 10669 bp in length * 64436 64444: gap of unknown length * 64445 74016: contig of 9573 bp in length * 74017 74105: gap of unknown length * 74105 83366: contig of 9262 bp in length * 83367 83454: gap of unknown length * 83455 92355: contig of 8901 bp in length * 92356 92443: gap of unknown length * 92444 100821: contig of 8378 bp in length * 100822 100909: gap of unknown length * 100910 107529: contig of 6620 bp in length * 107530 107617: gap of unknown length * 107618 114066: contig of 6449 bp in length * 114067 114154: gap of unknown length * 114155 118873: contig of 4719 bp in length * 118874 118961: gap of unknown length * 118962 123620: contig of 4658 bp in length * 123620 123707: gap of unknown length * 123708 128240: contig of 4533 bp in length * 128241 128328: gap of unknown length * 128329 128362: contig of 4354 bp in length * 128363 132770: gap of unknown length

FEATURES		* 132771 136341: contig of 3571 bp in length		* 136342 136429: gap of unknown length		* 136430 138070: contig of 1641 bp in length.	
source		Location/Qualifiers		1..138070			
		/organism="Mus musculus"		/mol_type="genomic DNA"		/db_xref="taxon:10090"	
		/clone="RP23-121P10"					
BASE COUNT	35731 a 33657 c 33303 g 33954 t	1425 others					
ORIGIN							
Query Match	66.9% Score 976.6; DB 2; Length 138070;						
Best Local Similarity	83.5% Pred No. 9,2e-209;						
Matches 1263; Conservative	0; Mismatches 184; Indels 66; Gaps 11;						
Db	10 GAGAGAGACAGTCCCTCGGAGCCCCCGCTTGCTGATGTGGCCCGT-GGCACAGGACAGACCCCG 68						
Qy	111713 GCGTGGCCGAGCCCTTAGGGCCCTCCGCTGATGTGGCCCGTGGGTGACAGGACAGACCCG 111654						
Db	69 GCAGGACAGCTCTCTGTGTCGGGACAGACGATTAAGCGTGGCCAGGGGACACAGATT--- 125						
Qy	111653 CCAGGACAGCTCTCTGCGGAGGACAGACGATTAAGCGGCCAGGGACACAGATTGG 111594						
Db	126 -AGCAGCTCAGAAAGTCCCTCTGGGTCTACCACTGCA-CAGAGGCGGAGGACCCCTCCG 183						
Qy	111593 ATGCAGCTCAGAAATCCCTCTGGGTCTACCTGACAGAGTGTGATGATCTCTCCG 111534						
Db	184 AGCTCTTGTGCTCTCCAGACGCAATTTACTCCAGGCGGCGCTCAGCTCAGTCACGA 243						
Qy	111533 AGCTTTTATGACACTCCAGACGCAATTTACTCCAGGCGGCGCTCAGTTTATGACG 111474						
Db	244 AACTTCGAGCGAGAGAGGGGTTCACTATCCACCGCTGCTTACTTGACC----- 296						
Qy	111473 AACTTCGAGCGAGAGAGGGGTTCACTATCCACCGCTGCTTACTTGAGTGGAG-AACTGAG 111414						
Db	297 -----ACCCGAGCTCTGTTCTTTTGGAGCCGAGATPACTAG 336						
Qy	111413 CTGCTGCTTGTCTACTGACTGACTCTGCTGCTCTTATTTTGTGAGTGGAG-AACTGAG 111355						
Db	337 TPACTATTAGAACTCTCCAAAGGTGAGAAGGGGAGT-----GGTGGCGTAC 386						
Qy	111354 TPACTATTGGAATCTCCAAAGGTGAGTGAAGGGGCGCGGGGTGTGTGTGGGGATAC 111295						
Db	387 TCTAGTCCCGCGTGAAGTGAAGTCTTGAAGTCAAGACTGTCAACCC--CTTCCATTTT 444						
Qy	111294 TCTGTCTCCCGTGAAGTGAAGTCTTGAAGTCAAGACTGTCAACCCCTTCCATTTT 111235						
Db	445 TTCCCACTCTGAGTGAAGTGGGCTCAATCCCTTGAATGGCGCCACATCCAAAGTGTCCCAAG 504						
Qy	111234 TTCCCACTCTGAGTGAAGTGGGCTCAATCCCTTGAATGGCGCTCAATCCAAAGTGTCCCAAG 111175						
Db	505 AGACCCAGCAACCTTTTCCCGAGCTCGAGCCAGCAAGTGTCAATTTCCACCC 564						
Qy	111174 AGACCAACCACTTTTCCCGAGCTCGAGCCAGCAAGTGTCAATTTCCACCC 111115						
Db	565 CACCTAGCCCACTCTGTTACCGAGGACTGCTCCAGCAAGAGCAAGTGAATGCGGAG 624						
Qy	111114 CACCTAGCCCACTCTCATTAAGGACTGCTCCAGCAAGAGTGAATGCGGAG 111055						
Db	625 GGAACATCGAGAAAGTCCGTGGCGCGCGCGGAGGGGCAACAGGCCCAAGAGGAGTTGG 684						
Qy	111054 GGAACATCGAGAAAGTCCGTGGCGCGCGCGGAGGGGCAACAGGCCCAAGAGGAGTTGG 110995						
Db	685 CACTAGCAAGAGGAGGAGCGGCGGCAAGAGGCAACGACCGGAGCGCAACCGCA 744						
Qy	110994 CACTAGCAAGAGGAGGAGCGGCGGCAAGAGGCAACGATTCGAGCGCAATTCGCA 110933						
Db	745 TCACAACTTAATCTCCGCTGTGATGCGCTGCGCGGTCTCTGCCACCTTCCCGATG 804						
Qy	110934 TCACAACTTAATCTCCGCTGTGATGCGCTGCGCGGTCTCTGCCACCTTCCCGATG 110875						
Db	805 ACGCAACTTAACAAAGATCGAGACCTTCGCTTCCGCCCAACATTAATTTGGGCACTGA 864						

Db 110874 ACGCCAACTTACAAAGATCAGACCTTCGCTTCCCAACAATCACTGGGACATCGA 110815
Qy 865 CTCAGAGCTGGGATAGGAGACACAGCTTCTAGGCCCCGAGCCCTGGCCCTGTG 924
Db 110814 CTCAGAGCTGGGATAGGAGACACAGCTTCTAGGCCCCGAGCCCTGGCCCTGTG 110755
Qy 925 GGAAGCTGGGAAGCCCGGAGGAGGAGCTCAGCGGAGCTGAGGCTCTACTACTCCAG 984
Db 110754 GAGAGCTGGGAGCCCGGAGGAGGAGCTCAGCGGAGCTGAGGCTCTACTACTCCAG 110695
Qy 985 TTTCCCAAGCTGGTATGCTGAGCCCAAGCTCTATTGAGAGAGTTCCCTGGCTGCAGG 1044
Db 110694 TCTCCCAAGCTGGTATGCTGAGCCCAAGCTCTATTGAGAGAGTTCCCTGGCTGCAGG 110635
Qy 1045 TGCCAGCTCCCAAGCTGCTGCTGCTCCGAGGAGCCCTGCTGCTCAGACTTCTGTGA 1104
Db 110634 TGCCAGCTCCCAAGCTGCTGCTGCTCCGAGGAGCCCTGCTGCTCAGACTTCTGTGA 110575
Qy 1105 GGGCCCAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1164
Db 110574 GAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110515
Qy 1165 GAATGAGAGTATGAGGAGCACTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1224
Db 110514 GAGAGTGGAGTATGAGGAGCTCTGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110455
Qy 1225 TCCCT-----GATTAAACAGAGTTCGCAAGTTCTGCTGCTGCTGCTGCTGCTGCT 1271
Db 110454 TCCCTAGCTCTCTCAAGGCTTAACTAGGCTTCTCACTGCTGCTGCTGCTGCTGCTGCT 110397
Qy 1272 ACAAGATATTTGAGGCTGATCTCTCTTAACTCTCTGAGTGGGACCTCAATCTG 1331
Db 110396 ACAGAGACATTTGGGAGCTGCTCTCTTAACTCTCTGAGTGGGACCTCAATCTG 110337
Qy 1332 CCGCTCAAGAGAGAGAGCCGCTGAGCACTAATATGTTGGAGACTCCCACTCTGCTG 1391
Db 110336 TCGCTCAAGAGATTTGAGATGATGAGCACTACTAGTT--GGAGACTCCCACTCTGCTG 110278
Qy 1392 TGACTCCGCTCTTTTAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1447
Db 110277 TGAGTCTGCTCTCAATTAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 110218
Qy 1448 CCTAATCCAGTGT 1460
Db 110217 CCTAATCCAGTGT 110205

RESULT 7
AF364300 1861 bp DNA linear ROD 03-MAY-2001
LOCUS AF364300
DEFINITION Mus musculus neurogenin 3 gene, complete cds.
ACCESSION AF364300.1 GI:13937128
VERSION AF364300.1
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Lee,J., Smith,S., Matada,H., Lin,J., Scheel,D., Wang,J., Mimitra,R.
and German,M.
Regulation of the pancreatic pro-endocrine gene neurogenin3
Diabetes (2001) In press
2 (bases 1 to 1861)
Schwartzgebel,V. and German,M.
Direct Submission
Submitted (26-MAR-2001) Hormone Research Institute, University of
California San Francisco, 513 Parnassus Ave, HSW1090, San
Francisco, CA 94145-0534, USA
Location/Qualifiers
1..1861
FEATURES
Source
/organism="Mus musculus"

Query Match 61.7%; Score 900.6; DB 10; Length 1861;
Best Local Similarity 86.2%; Pred. No. 1.7e-191;
Matches 1094; Conservative 0; Mismatches 129; Indels 46; Gaps 7;
CDS
mRNA
/mol_type="genomic DNA"
/db_xref="taxon:10090"
<1093..>1737
/product="neurogenin 3"
1093..1737
/codon_start=1
/product="neurogenin 3"
/protein_id="AAK5058.1"
/db_xref="GI:13937129"
/translation="MAPHPDALTIQVSPETPOEPFPGASPDHEVLSNSTPSPPLIR
DCEBAYDQRCSTRLKRRGRNRKSTALSKORSRKRNDRNDNRNINA
LDLRGVLPFPDDAKTET
PFGDSNWDWSTISVPSVQANLSPTALEPFLQVSPSPSYLLPGLVPSDFL"
BASE COUNT 397 a 560 c 537 g 367 t
ORIGIN
4 GGTAGCAGAGAGAGACAGTCCCTGGGAGCCCGCTTCTGATTTGGCCGTCAGACAGCA 63
594 GGGCTCGGTGGCCAGCCCTGGGAGCCCTCCGCTAATTTGGCCGTCAGACAGCA 653
64 GCCCGAGAGGACGCTCCTGCTC--CGGAGAGAGAGATTAAGCTGCGAGGAGACAGC 122
654 GCCCGAGAGGACGCTCCTGCGGGGAGAGAGAGATTAAGCTGCGAGGAGACAGC 713
Qy 123 ATT---AGAGCTCAGAGTCCCTTGGGCTTCAACACTGCA--CAGAGCCGAGAACCC 177
Db 714 ACTTGACATGACGCTCAAAAATCCCTGAGGCTCTCATCATGACAGAGTGTGAGTACT 773
Qy 178 CCTCGAGCTCTTTGCTGCTCCTCAGAGCAATTTACTCCAGGAGGAGGAGGCTGAGCT 237
Db 774 CTTGAGCTTTTCTTACAGCTTCCAGACGATTTACTCCAGGAGGAGGAGGCTGAGCT 833
Qy 238 CAGCAAACTTCAAGAGAGAGAGGAGGATTCAGCTATCAACCGCTGCTTCACTGAC-- 296
Db 834 TAGCAGAACTTCAAGAGAGAGAGAGGAGGAGTACAGCTGCTGCTTGAACAAGACC 893
Qy 297 -----ACCGAGCTCTGCTTTTGAAGCCGAGATA 330
Db 894 TATCAGTGTGCTTGTCACTGACAGTCACTGCTGCTGCTCTCTAATCTTTGAGTCGGAG--A 952
Qy 331 ACTAGTAACTTTAGAACTCCCAAGGAGTGAAGAGGAGGAGT-----GG 380
Db 953 ACTAGTAACTTTAGAACTCCCAAGGAGTGAAGAGGAGGAGGAGGAGTGTGTGGG 1012
Qy 381 GCGTACTAGTCCCGGTGAGTGAAGCTTAACTGAGAGAGT--TCACACCCCTTC 438
Db 1013 GGAATCTGTGCTCCCGGTGAGTGAAGCTTAACTGAGAGAGTGTGACACACACACCTTC 1072
Qy 439 CATTTTTCCTCACTGAGATGAGGCTTATCCCTTGATGCGCCACATCCAGTGT 498
Db 1073 CATTTTTCCTCACTGAGATGAGGCTTATCCCTTGATGCGCTCAACATCCAGTGT 1132
Qy 499 CCCAAGAGCCAGCAACCTTTCCCGAGAGCTTGGAGACAGAGTGTCTAGTCCCAATT 558
Db 1133 CCCAAGAGCAACAACCTTTCCCGAGAGCTTGGAGACAGAGTGTCTAGTCCCAATT 1192
Qy 559 CCACCCACCTAGGCCCCACTCTGTAACGAGAGAGTCTCCGAGAGAGAGAGTACT 618
Db 1193 CCACCCACCTAGGCCCCACTCTGTAACGAGAGAGTCTCCGAGAGAGAGTGTACT 1252
Qy 619 GCCGAGGAGATCGAGAGAGCTCGTGCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 678
Db 1253 GCCGAGGAGATCGAGAGAGCTCGGCGCCGAGCGGAGGAGGAGGAGGAGGAGGAGG 1312
Qy 679 AGTTGACATGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 738
Db 1313 AGTTGACATGAGAGAGAGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1372
Qy 739 ACCGATGACAACTTAATCTCGGCTGATGAGTGTGCGGAGTGTCTGCTCCACTTCC 798

Db	1373	ATCGCATGCAAAACCTTCAACTCTGCGCTGAGATGCGCTCGCGGATCTCTGCCACCTTCC	1432
Qy	799	CGAGTGAAGCCAACTTTCACAAAGATCGAGACCCCTGCGCTTCGCCACAACTACATTGGG	858
Db	1433	CGAGTGAAGCCAACTTTCACAAAGATCGAGACCCCTGCGCTTCGCCACAACTACATTGGG	1492
Qy	859	CACGTACTCAGACCGCTGCGCATATGCGGACACACGTTCTTACGGCCCGCAGCCCTCTGTC	918
Db	1493	CACGTACTCAGACCGCTGCGCATATGCGGACACACGTTCTTATGCGCCCGAGCCCTCTGTC	1552
Qy	919	CCTGTGGGGAGCTGGGAAGCCCGGAGGGGGGCTCCAGGGGCGACTGGGGCTCTATCTACT	978
Db	1553	CCTGTGGGGAGCTGGGAAGCCCGGAGGGGGGCTCCAGGGGCGACTGGGGCTCTATCTACT	1612
Qy	979	CCCCAGTTTCCCAAGCTGTAGCCTCGAGCCGCCACAGCCTCTCATTTGAGAGAGTTCCTGACC	1038
Db	1613	CCCCAGTTTCCCAAGCGGGTAACTCGAGCCGCCACAGCCTCTCATTTGAGAGAGTTCCTGACC	1672
Qy	1039	TGCAGGTGCCAGCTCCCATCTCTGTCTGCTCCGGGACACCTGTGTTCTCAGACTTCT	1098
Db	1673	TGCAGGTGCCAGCTCCCATCTCTGTCTGCTCCGGGACACCTGTGTTCTCAGACTTCT	1732
Qy	1099	TGTGAAGGGCCCAACAGAGCCCTGGGGCGGTGGGGCGCTGGCAAGAGAGGAGTCAAG	1156
Db	1733	TGTGAAGAGACTGTTCTGCTCTGTGGGTGGGTGCTATGTGAAGAGAGGAGGACCAAG	1792
Qy	1159	CTGTCTGAATGAGAGTATGAGAGCACTCGAGCATCTCGCCCTTGTGCTTTCATTA	1218
Db	1793	CGGTGTGAAGTGGGAGTATGAGAGCACTCGAGCATCTCGCCCTTGTGCTTTCATTA	1852
Qy	1219	GTCAGCTCC 1227	
Db	1853	CTGTGATCC 1861	

RESULT 8	MMATH4B	5567 bp	DNA	linear	ROD 25-OCT-2000
LOCUS	MMATH4B				
DEFINITION	M.musculus MATH4B gene.				
ACCESSION	Y09167				
VERSION	Y09167.2	GI:11065737			
KEYWORDS	bHLH; mammalian atonal homolog 4B; MATH4B; neurogenin 3.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
REFERENCE	1	Cau, F., Gradwohl, G., Fode, C. and Guillemot, F.			
AUTHORS	2	Mashi activates a cascade of bHLH regulators in olfactory neuron progenitors			
JOURNAL	3	Development 124 (8), 1611-1621 (1997)			
MEDLINE	97261963				
PUBMED	9108377				
REFERENCE	4	Jacquemin, P., Durvieux, S.M., Jensen, U., Godfraind, C., Gradwohl, G., Guillemot, F., Madсен, O.D., Carmeliet, P., Dewerchin, M., Collen, D., Rousseau, G.G. and Lemaigre, F.P.			
AUTHORS	5	Transcription factor hepatocyte nuclear factor 6 regulates pancreatic endocrine cell differentiation and controls expression of the proendocrine gene ngn3			
JOURNAL	6	Mol. Cell. Biol. 20 (12), 4445-4454 (2000)			
MEDLINE	20285449				
PUBMED	10825208				
REFERENCE	7	Gradwohl, G.J.			
AUTHORS	8	Direct Submission			
JOURNAL	9	Submitted (04-NOV-1996) G.J. Gradwohl, IGBMC, CNRS-INSERM-Universite Louis Pasteur, BP163, C.U. de Strasbourg, F-67404 ILLKIRCH cedex, FRANCE			
REMARK	10	Revised by [4]			
REFERENCE	11	4 (bases 1 to 5567)			
AUTHORS	12	Lemaigre, F.P.			

FEATURES	source	1..5567
COMMENT	location/Qualifiers	
FEATURES	1..5567	
misc_binding	/organism="Mus musculus"	
misc_binding	/mol_type="genomic DNA"	
misc_binding	/strain="SV129"	
misc_binding	/db_xref="taxon:10090"	
misc_binding	/clone="lambda GEM12 clone 19C"	
misc_binding	/cell_line="D3"	
misc_binding	/cell_type="embryonic stem cells"	
misc_binding	/clone_id="genomic library No.62, MboI partial digests cloned in BamHI"	
misc_binding	/dev_stage="embryos"	
misc_binding	/bound_moiety="HNF6"	
misc_binding	3692	
misc_binding	/bound_moiety="HNF6"	
misc_binding	4145..5567	
gene	/gene="MATH4B"	
gene	4145..4151	
gene	/gene="MATH4B"	
gene	/note="putative"	
gene	4923..5567	
gene	/gene="MATH4B"	
gene	/codon_start=1	
gene	/product="neurogenin 3"	
gene	/protein_id="CAI70366.1"	
gene	/db_xref="GI:1666088"	
gene	/db_xref="SWISS-PROT:P70661"	
gene	/translation="MAPHPDALTITQVSPETQEPFSPASDHEVLSSNSTPSPPLIPRDCSAYGVDCRGTSRKLRKRGGRNRKSEIALSKORSRKKANDERNMHNLGSLDARGVLPFPDDAKLTLETLRFANHYIWTALQTLRIDHSFYGEPPVPCGELGSPGGSNDMCSISYSPVSAQNLSPYASILEEPFGIVPSPEYLLPLALVFSDFL"	
BASE COUNT	1271 a 1549 c 1564 g 1183 t	
ORIGIN		
Query Match	56.2%; Score 819.8; DB 10; Length 5567;	
Best Local Similarity	87.1%; Pred. No. 2e-173;	
Matches	997; Conservative 0; Mismatches 102; Indels 46; Gaps 7;	
4	GGTAGCGAGAGAGAGAGTCCCTGTGGGCCCCGTTGCTGATTGGCCCGTGGCACAAGGACAGA 63	
4424	GAGGCTCGGCGTGGCGACAGCCCTGTGGGCCCCCTCCGCTGATTGGCCCGTGGTCAAGGACAGA 4483	
64	GCCCGGCGAGAGAGAGCTCCTGTGGT-CGGGCGAAGAGAGATTAAGCGTCCAGGGGACACAGC 122	
4484	GCCCGGCGAGAGAGAGCTCCTGTGGCCGGGGCGAAGAGATTAAGCGTCCAGGGGACACAGC 4543	
123	ATT-----AGCAGCTCAGAAAGTCCCTCTGGGTCTCAACCACTGCA-CAGAGGCGGAGACC 177	
4544	ACTTGCATGCAAGCTCAAGAAATCCCTCTGGGTCTCATCACTGCAAGAGTGTGAGTAACT 4603	
178	CTTCGAGCTTCTTTGCTGCTCCAGACGCAATTTACTCAGGGGAGGGCGCTGCACT 237	
4604	CCTGGAGCTTTTCTACAGACTTCCAGAGCGCAATTTACTCAGGGGAGGGCGCTGCACTT 4663	
238	CAGCAAAATTTGAAAGAGAGAGAGGGTTAGCATATCAACCGCTGTGACTGTGAGC- 296	
4664	TAGCAGAACTTCAGAGGGAGAGAGAGGCTAGCATATCATCTGCTGTGACATGACCCC 4722	
297	-----ACCCGAGCTCTGTCTTTGAGCCCGGAGTA 330	
4724	TATCCACTGCTGCTTGTACTGACTGACTGACTGCTGTCTATTTCTTTGAGTGGGAG-A 4783	
331	ACTAGTAACTTTAGAACTCCAAAGGTTAGAAAGAGGGAGTG-----GGTGG 380	
4783	ACTAGTAACTTTGAAACTCCAAAGGTTAGAAAGAGGGGCGCGGGGTGTGTGTGGG 4844	
381	GCGTACTTAACTCCCGCGGTGAGATGACCTCTTAAGTCAGAGACTGCAACCCC--CTTC 438	


```

Db      4843  GGAATCTGTGCCCCCGTGCAGTGAAGCTTAAGCAAGGGCTGGACACACACACCTTC 4902
Qy      4439  CATTTTTCCTCCAACTCAGATGGGCGCTCACTCTGGATGCGGCCACCAATCCAAAGT 498
Db      4903  CATTTTTCCTCCAACTCAGATGGGCGCTCACTCTGGATGCGGCCACCAATCCAAAGT 4962
Qy      499   CCCAAGAGACCCAGCAACCTTTTCCCGAGGCTCCGAGCCAGCAAGTGTCTAGTTCCAA 558
Db      4963  CCCAGAGACACAAACACTTTTCCCGAGGCTCGGACCAAGATGTCTAGTTCCAA 5022
Qy      559   CCACCCCACTAGCCCACTCTCTGTAACGAGGGAATGCTCCGAGACCAAGAGAGTGA 618
Db      5023  CCACCCCACTAGCCCACTCTCTGTAACGAGGGAATGCTCCGAGACCAAGAGTGA 5082
Qy      619   GCCAGGAGACATTCGAGGAGCTCCGTCGCGCGCGGAGGAGGAGCAAGGCCCAAGAG 678
Db      5083  GCCAGGAGACCTTCGAGGAGCTCCGTCGCGCGCGGAGGAGGAGCAAGGCCCAAGAG 5142
Qy      679   AGTTGGCACTAGACGAGAGCGACGAGCCGCGCAAGAGGCAACGAGCGGAGCGCA 738
Db      5143  AGTTGGCACTAGACGAGAGCGAGAGCGCGCAAGAGGCAATGATCGGAGCGCA 5202
Qy      739   ACCGATCAACAACCTTAATCTCCGCTGAGATGCGCTGCGCGGTCTGCTGCCACCTTC 798
Db      5203  ATCCGATCAACAACCTTAATCTCCGCTGAGATGCGCTGCGCGGTCTGCTGCCACCTTC 5262
Qy      799   CGGATGAGCGCCAACTTACAAAGATCGAGACCCCTGCGCTGCGCCCAACATATTTGG 858
Db      5263  CGGATGAGCGCCAACTTACAAAGATCGAGACCCCTGCGCTGCGCCCAACATATTTGG 5322
Qy      859   CACTGACTCAGACGCTGGCGATAGCGGACCAACAGCTTCTACGCGCCCGAGCCCTG 918
Db      5323  CACTGACTCAGACGCTGGCGATAGCGGACCAACAGCTTCTATGCGCCCGAGCCCTG 5382
Qy      919   CCTGTGGAGCTGGGAAAGCCCGGAGGCGGCTCCAGCGGCGACTGGGCGCTCTATCTA 978
Db      5383  CCTGTGGAGCTGGGAGAGCCCGGAGGCTCCAGCGGCGACTGGGCGCTCTATCTA 5442
Qy      979   CCCCACTTCCCAACTGCTGAGCTGAGCCCAACAGCTCTCATTTGAGAGAGTTCTTGG 1038
Db      5443  CCCCACTTCCCAACTGCTGAGCTGAGCCCAACAGCTCTCATTTGAGAGAGTTCTTGG 5502
Qy      1039  TGCAGGTGCCAGCTCCCATCTGCTCTCCGCGGACCCCTGGTGTCTTCAACTCT 1098
Db      5503  TGCAGGTGCCAGCTCCCATCTGCTCTCCGCGGACCTGCTGTCTTCAACTCT 5562
Qy      1099  TGTGA 1103
Db      5563  TGTGA 5567

RESULT 9
AX698801 861 bp DNA linear PAT 02-APR-2003
LOCUS Sequence 7 from Patent WO02086107.
DEFINITION AX698801
ACCESSION AX698801
VERSION AX698801.1 GI:29499589
KEYWORDS
SOURCE
ORGANISM Mus musculus (house mouse)
MUS musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 Wobus, A.M., St-Onge, L., Blyszczuk, P. and Hoffmann, U.
A method for differentiating stem cells into insulin-producing
cells
JOURNAL Patent: WO 02086107-A 7 31-OCT-2002;
DEVELOPER Deutscher Forschungszentrum fuer Entwicklungstechnische Forschung
(TUEBINGEN)
AUTHORS (DE) ; INSTITUT FUER PFLANZENGENETIK UND KULTURPFLANZENFORSCHUNG
(TUEBINGEN)
FEATURES
source 1..861
location/Qualifiers

```

```

CDS
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
160..804
/feature="unlabeled protein product"
/codon_start=1
/protein_id="CAD87442.1"
/db_xref="GI:29499590"
/translation="MAPK1DALT1QVSPETQPPGASDHEVLSSNTPSPILIPR
DSEAEVDRGRTSRKLRRGRNRKSRSLASKORSRKRNDRNRNMLNSA
LDLRGLVPTPPDPAKLTKEITLFAFNRNYIALQTLRIADHSRYGEPPIPCBELGS
PGGSNDWMSIYSPVQAGNLSPTALBEPFGLQVPSPSYILPLGALVDFPL"
BASE COUNT 182 a 274 c 250 g 155 t
ORIGIN
Query Match 46.7% Score 682.4; DB 6; Length 861;
Best Local Similarity 89.7% Pred. No. 1.6e-142;
Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;
Qy 312 TTCTTTGAGCCCGAGATTAAGTAACTTTAGAACTCCAAAGGATGAGAGGG 371
Db 2 TTCTTTGAGTGGGAG-ACCTAGTAACTATGGAACCTCAAAGGATGAGAGGG 60
Qy 372 AGTGGT-----GGCGTACTAGTCCCGCGTGAAGTGAAGTGAAGTGAAG 421
Db 61 CGCGCGGGGTGTGTGGGGATTAAGTCTGCTCCCGCGAGTGAAGTGAAGTGAAG 120
Qy 422 CTG--TCAACCCCTTCCATTTTCCCAACCTCGAGATGGGCGCTATCCCTGGAT 479
Db 121 CTGGACACACACACTTCCATTTTCCCAACCGAGATGGGCGCTATCCCTGGAT 180
Qy 480 GCGCCACCATCAAGTGTCCCAAGAGACCGAGAACCTTTCCCGAGCTCGACAC 539
Db 181 GCGCTCACATCAAGTGTCCCAAGAGACCGAGAACCTTTCCCGAGCTCGACAC 240
Qy 540 GAAGTCTCAATTCATTCACCCCACTAGCCCACTCTGTAACGAGGATGCTCC 599
Db 241 GAAGTCTCAATTCATTCACCCCACTAGCCCACTCTGTAACGAGGATGCTCC 300
Qy 600 GAACGAGAGAGGTGATGCTGGAGGAGCATTCGAGAGAGTCCGCGCGCGCGAGAG 659
Db 301 GAACGAGAGAGGTGATGCTGGAGGAGCATTCGAGAGAGTCCGCGCGCGAGAG 360
Qy 660 CGCAACAGGCGCCAGAGGAGTGGACCTGAGCAAGAGAGAGAGAGAGAGAG 719
Db 361 CGCAACAGGCGCCAGAGGAGTGGACCTGAGCAAGAGAGAGAGAGAGAGAG 420
Qy 720 GCCAAGAGCGGAGCGCAACCGATGACCAACCTTAATCTCGCGCTGATGCGTGGC 779
Db 421 GCCAATGATGGGAGCGCAATCGATGACCAACCTTAATCTCGCGCTGATGCGTGGC 480
Qy 780 GGTGTCTGCGCCACTTCCGAGTGAAGCCCAACTTACAAAGATGAGAGCTTGCCTTC 839
Db 481 GGTGTCTGCGCCACTTCCGAGTGAAGCCCAACTTACAAAGATGAGAGCTTGCCTTC 540
Qy 840 GCCCAACATCAATTTGGGAGCTGACTGAGCGTGGAGCATAGGAGGACACAGCTTAC 899
Db 541 GCCCAACATCAATTTGGGAGCTGACTGAGCGTGGAGCATAGGAGGACACAGCTTAC 600
Qy 900 GGGCCGAGCCCTGTGCTCTGTGGAGAGCTGGAGAGCCCGGAGGAGGCTCCAGCGC 959
Db 601 GGGCCGAGCCCTGTGCTCTGTGGAGAGCTGGAGAGCCCGGAGGAGGCTCCAGCGC 660
Qy 960 GACTGGGCTTATCTACTCCCAAGTTTCCCAAGTGTAGCTGAGCCCAAGCTTCA 1019
Db 661 GACTGGGCTTATCTACTCCCAAGTGTAGCTGAGCCCAAGCTTCA 720
Qy 1020 TTGAGAGAGTCCCTGGCTGAGAGTCCCAAGTGTAGCTGAGCTGAGCTGAGCTGAG 1079
Db 721 TTGAGAGAGTCCCTGGCTGAGAGTCCCAAGTGTAGCTGAGCTGAGCTGAGCTGAG 780
Qy 1080 CTGTGTCTCAGACTTCTTGTGAAGGCGCCAAAGGCGCTGGGCGTGGCA 1139

```

Db 781 CTGGTCTCTAGACTTCTTGGAAGAGACCTGTGCTGTGGTGTGGTGTCTAGTG 840

Qy 1140 GAAGGAGGAGGAGTCAAGC 1159

Db 841 GAAAGGAGGAGGAGCAGAC 860

RESULT 10

LOCUS MMU76208 861 bp DNA linear ROD 05-FEB-1997

DEFINITION Mus musculus neurogenin 3 (ngn3) gene, complete cds.

ACCESSION U76208

VERSION U76208.1 GI:1815654

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 861)

AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.

TITLE neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS

JOURNAL Mol. Cell. Neurosci. 8 (4), 221-241 (1996)

MEDLINE 97153565

PUBMED 9000438

REFERENCE 2 (bases 1 to 861)

AUTHORS Sommer, L., Ma, Q. and Anderson, D.J.

TITLE Direct Submission

JOURNAL Submitted (24-OCT-1996) Biology 216-76, California Institute of Technology, Howard Hughes Medical Institute, Pasadena, CA 91125, USA

COMMENT On Feb 5, 1997 this sequence version replaced gi:166911.

FEATURES

source location/Qualifiers

1..861

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

160..804

/gene="ngn3"

160..804

/gene="ngn3"

/function="BHLH transcription factor"

/note="NGN3"

/codon_start=1

/product="neurogenin 3"

/protein_id="AAC53029.1"

/translation="MAPHPIDALTIQVSPETQPPGASDHEVLSSNSTPSPFLIPR DCEAEVDGCRGTSRKLRRRGRRNRPSESLASRORSRRKADRRNRHNNLSA LDARGVLPFRPDPAKLTETLRLRAHYIYALITOTLRADSPFGPRPVPVCGELSG PGGSGNDGSIYSFVSQGNLSPTASLEFPDLPSPSYLFGALVSDFL"

BASE COUNT 182 a 274 c 250 g 155 t

ORIGIN

Query Match 46.7%; Score 682.4; DB 10; Length 861;

Best Local Similarity 89.7%; Pred. No. 1.6e-142;

Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;

Qy 312 TTTCTTTGAGCCCGGAGTAAGTAACTTGAAGAACCTCCAAAGGTAAAGAGGG 371

Db 2 TTTCTTTGAGTGGGAG- AACTAGGTAAACATTCGAAATCTCCAAAGGTGATAGGGG 60

Qy 372 AGTGGT-----GGGCGTACTTGAATCCCGGTGAGTGAATCTTAAAGTCAGAGA 421

Db 61 CGCGCGGGGTGTGTGNGGGGATACCTGTGCTCCCGGTGAGTGAATCTTAAAGTCAGAG 120

Qy 422 CTG--TCACACCCCTCTTCAATTTTCCCAACTCAGATGCGGCTCATCCCTTGAT 479

Db 121 CTGGACACACACACCTTCAATTTTCCCAACCGAGATGCGGCTCATCCCTTGAT 180

Qy 480 GGGCCACCATCAAGTGTCCCAAGAGACCCGAGCACTTTCCGGAGCTCGACAC 539

Db 181 GCGTCACCATTCAGATGTCCCAAGAGACACACAAACCTTTTCCGGAGCTCGAGCAC 240

Qy 540 GAAGTGTCAAGTTCATTCACCCACCTGACCCCACTCTGTACGAGGAGCTGTCC 599

Db 241 GAAGTGTCAAGTTCATTCACCCACCTGACCCCACTCTGTACGAGGAGCTGTCC 300

Qy 600 GAAGCAGAGAGGTGACTGCGGAGAGCATGAGAAAGCTCCGTGCGGCGGAGAGG 659

Db 301 GAAGCAGAGAGGTGACTGCGGAGAGCATGAGAAAGCTCCGTGCGGCGGAGAGG 360

Qy 660 CGCAACAGGCGCCAGAGGAGTGGCACTGAGCAAGCAGCAGCAAGCCGCGCAAGAG 719

Db 361 CGCAACAGGCGCCAGAGGAGTGGCACTGAGCAAGCAGCAGCAAGGCGCGCAAGAG 420

Qy 720 GCCAAGCAGCGGAGCGCAACCGCATGACCAACCTTAACTCGCGCTGATGCGTGGC 779

Db 421 GCCAATGATCGGAGCGCAATCGCATGACCAACCTTAACTCGCGCTGATGCGTGGC 480

Qy 780 GGTGTCTGCGCCACCTTCCCGATGACCGCAACTTACCAAGATGAGACCTGCGCTTC 839

Db 481 GGTGTCTGCGCCACCTTCCCGATGACCGCAACTTACCAAGATGAGACCTGCGCTTC 540

Qy 840 GCCCAACTACATTTGGGCACTGACTAGACGCTGCGCATAGCGGAGCAGCAAGCTTAC 899

Db 541 GCCCAACTACATTTGGGCACTGACTAGACGCTGCGCATAGCGGAGCAGCAAGCTTAC 600

Qy 900 GGGCCGAGCGCCCTGTGCTGCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 959

Db 601 GGGCCGAGCGCCCTGTGCTGCGGAGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGG 660

Qy 960 GACTGGGCGCTATCTACTTCCCAAGTTTCCCAAGCTGTAGCTGAGGCCAGCAAGCTCA 1019

Db 661 GACTGGGCGCTATCTACTTCCCAAGTTTCCCAAGCTGTAGCTGAGGCCAGCAAGCTCA 720

Qy 1020 TTGAGAGATTCCTGCTGAGAGTGGCCAGCTGCCATCTGTCTCTCCGCGGAC 1079

Db 721 TTGAGAGATTCCTGCTGAGAGTGGCCAGCTGCCATCTGTCTCTCCGCGGAC 780

Qy 1080 CTGGTCTCTAGACTTCTTGGAAGAGGCGCAACAGGCGCTGGGCGGTGGCGCTGCA 1139

Db 781 CTGGTCTCTAGACTTCTTGGAAGAGGCGCAACAGGCGCTGGGCGGTGGCGCTGCA 840

Qy 1140 GAAGGAGGAGGAGTCAAGC 1159

Db 841 GAAAGGAGGAGGAGCAGAC 860

RESULT 11

AF234829 5340 bp DNA linear PRI 19-OCT-2001

LOCUS Homo sapiens neurogenin 3 gene, complete cds.

DEFINITION AF234829

ACCESSION AF234829.1 GI:13183002

VERSION

KEYWORDS

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 5340)

AUTHORS Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euteleostomi; Rodentia; Sciurognathi; Muridae; Murinae; Mus. del Bopque-Plata, L., Lin, J., Horikawa, Y., Schwarz, P.E., Cox, N.J., Iwasaki, N., Ogata, M., Iwamoto, Y., German, M.S. and Bell, G.I.

TITLE Mutations in the coding region of the neurogenin 3 gene (NEUROG3) are not a common cause of maturity-onset diabetes of the young in Japanese subjects

JOURNAL Diabetes 50 (3), 694-696 (2001)

MEDLINE 11246894

PUBMED 11240923

REFERENCE 2 (bases 1 to 5340)

AUTHORS Lin, J. and German, M.

TITLE Direct Submission

JOURNAL Submitted (15-FEB-2000) Hormone Research Institute, University of California San Francisco, 513 Parnassus Ave., San Francisco, CA 94143-0534, USA

FEATURES
source Location/Qualifiers
1..5340
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
3022..3666
/product="neurogenin 3"
/codon_start=1
/product="neurogenin 3"
/protein_id="AAK15022.1"
/db_xref="GI:13183003"
/translation="MTPOPSSAPVQVTRERESFPRASEDEVCTPSAPSPRTG
NCAEEBEGCRGAPKLRARGRSRPKSELALSKORSRKANDRENNMINSLA
LDALRGVLPTEPDADKLKRIETLRAPHNYIWALOTLEIADHSIYALPPAPHCSEIG
SPGSGPDMGSLVSPVSGASLSPASISLERPGLIGATISNACLSRGLAFSDFL"

CDS
1215 a 1500 c 1514 g 1111 t

BASE COUNT
1215 a 1500 c 1514 g 1111 t

ORIGIN

Query Match 35.4%; Score 517.4; DB 9; Length 5340;
Best Local Similarity 67.4%; Pred. No. 1.3e-105;
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

20 GTCCCTGGCCCGCGTCTGATTGGCCCGTGGCAGAGCAGCCCGGAGGAGCGCT 79
|||
2535 GGCCCGAGGCGCCGCGCTGATTGGCGGTTGGCGGAGCAGCCGGGAGGAGCGCT 2534
|||
80 CCTGGTCCGGGAGAGCAGATTAAGCGTGGCCAGGGGAGACACGATTACAGTCAAGT 139
|||
2595 CCTGGCCGGGCGAAGAGATTAAGCGTGGCCAGGGGAGACACGATTACAGTCAAGT 2654
|||
140 CCTCTGGGCTCAACCATGCG-----ACGAGAGCGGAGGAGCCCGCTC 181
|||
2655 TCCCTGGGCTCAACCATGCGCGCGCTCGAGAGAGCGTGAAGAGCGCTCGAGCCCTT 2714
|||
182 CGAGCTCTTTGCTGCTCCAGACGCAATTACT-CCAGGCGAGGCGCGCTGCACTCAG 240
|||
2715 CTCTCTTTCTTTCTCTTTGGGGCTGGGGCAACTCCAGCGGGGCGCTGCACTCAG 2774
|||
241 CAAATCTTGAAGCAGCAG-----AGGGGTTCAGTATCCACCGCTGCT 285
|||
2775 CTGAATCTGGGAGCAGAGCGCGCTGAGCTCCCGAGCGCTCTCTCTGATGCTTC 2834
|||
286 TGACTCTGACCAACCGGAGCTCTCTGTTCTTTTGAAGCGGAGTAACTAGGTAACATTTA 345
|||
2835 TATTCTTTGGCGCGGTGAAGAGTAATTTTGGAGCGCTCCGAGGAGCGGAGGAGGAA 2894
|||
346 GGAAC-CTCAAGAGGTGAAGAGGAGGTGGGTGGGTACTAGTCCCGGTGAGT 404
|||
2895 AGAGGATCTCTGACCCAGCGGGGCTGGAGATGCTGTTTTTTTCCCACT 2954
|||
405 GACCTCTAAGTCAGAGACTG-----TACACCCCGCTTCATTTTCCCA 451
|||
2955 AGCCTCGGAATCGCGAGCTGGCGCGCTGAAGACTCAACCTTCCCTCTGACCCCG 3014
|||
452 CCTGAGATGGGCGCTCATCCCTTGGAGCGGCCCATTCAGAGTCCCAAGAGAGACCA 511
|||
3015 CCGTAGATGAAGCTCAACCTCGGGTGGCGCCACTGTCAGAGTACCGGTGAGACGGA 3074
|||
512 GCAACCTTTCCCGAGGCTCGAGACCAAGAGTCTCAATTCACCCCACTAG 571
|||
3075 GGGGCTTTCCCGAGGCTCGAGAGAGAGTAGTACCTGCCCACTCGGCCCGCCAG 3134
|||
572 CCCCACTCTCGTACCGAGGAGTCTCGGAAGCAGAGCAGTGACTCCGAGGAGATC 631
|||
3135 CCCCACTCGACACCGGGGAACTGCGCAGAGCGGAGAGAGGAGGCTCCAGAGGGGCC 3194
|||
632 GAGGAGCTCGGTGGCGGCGGAGGCGGAGGCGCAAGGCGCCAAAGCAGGATTGCACTAG 691
|||
3195 GAGGAGCTCGGGGAGCGCGGGGAGCGAGCGCGCTTAAGACGAGTTGCACTAG 3254
|||
692 CAAGCAGCAGAAACCGGCGCAGAGAGGCCAACGAGCGGAGCGCAACCGCATGCAAA 751
|||

Db 3255 CAAGCAGCAGACGAGTGGCGGAGAAAGAGCCCAACGACCGGAGCGCATGCAATGCAACAA 3314
|||
Oy 752 CCTTAATCTCGCGCGTGGATGGCGTGGCGGAGTCTCTGACCACTTCCCGATGACGCCAA 811
|||
Db 3315 CCTTAATCTCGCGCGTGGATGGCGTGGCGGAGTCTCTGACCACTTCCCGATGACGCCAA 3374
|||
Oy 812 ACTTACAAAGATGAGACCCCTGCGCTTCCGACACACTTATTTGGGCACTGACTCAGAC 871
|||
Db 3375 GCTCACAAGATGAGACCGCTGCGCTTCCGACCACTTATTTGGGCGGCTGACTCAAAAC 3434
|||
Oy 872 GCTGCGCATGAGCGGACCAACGCTTCTACGCGCCCGGACCCCGCTGCGCTGAGGAGCT 931
|||
Db 3435 GCTGCGCATGAGCGGACCAACGCTTCTACGCGCGCGGACCCCGGCGCGCACTCGGAGGA 3494
|||
Oy 932 GGAAGCGCGGAGAGGGGCTTCCAGCGCGGAGCTGAGGCGTCTATCTACTCCCACTTCCCA 991
|||
Db 3495 GCTGAGGAGCGGAGCGGCTTCCCGGAGGAGTGGGCTCCCTCTACTCCCACTTCCCA 3554
|||
Oy 992 AGCTGTAGCCTGAGCCCAACGCTTATGAGAGGAGTCCCTGCGCTGAGTCCCGAG 1051
|||
Db 3555 GCTGAGGAGCCTGAGTCCCGCGCGTCTGAGAGAGCAGACCCGGGCTGAGGAGGCGCAC 3614
|||
Oy 1052 CTCCCACTCTGCTGCTCCCGGAGCACCCTGAGTCTTCAAGTCTTTGTAAGAGGCCCA 1111
|||
Db 3615 CTCTTCCGCTGCTTGAAGCCAGGCACTGCGCTTCTCAAGTTTCTGTGAAGAGACT 3674
|||
Oy 1112 AAGAGCGCTGGGCGGAGGCGCTGAGCAAGAGGAGGAGTCAAGCTGTCTGAATG 1171
|||
Db 3675 GTCTGTGCTGGGCTGAGGCTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3734
|||
Oy 1172 AAGGTAGTGAAGCAGCTGAGCATC 1196
|||
Db 3735 AGGTTGGCCAGCGGCGGCGGCTC 3759
|||

RESULT 12
AL450311/c
LOCUS
DEFINITION
Human DNA sequence from clone RP11-343J3 on chromosome 10, complete
sequence.
AL450311
VERSION
AL450311.11 GI:14626972
HTG.
ACCESSION
AL450311.11
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 165110)
Howden, P.
TITLE
Direct Submission
JOURNAL
Submitted (12-JUL-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Jul 8, 2001 this sequence version replaced gi:14575291.
During difference assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em., EMBL, SW.,
SWISSPROT, Tr., TrEMBL, Wp., WormPep; information on the WormPep
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human

Chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr10>
Rp11-343J3 is from the library RPCT-11.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBACe3.6
This sequence is the entire insert of clone Rp11-343J3 The true left end of clone Rp11-242G20 is at 139955 in this sequence. The true right end of clone Rp11-404C6 is at 6588 in this sequence.

FEATURES

source
1. .165110
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-343J3"
/clone_11b="RPCT-11.2"
7. .147
/note="THE1C repeat: matches 2. .142 of consensus"
repeat_region
151. .293
/note="MIR repeat: matches 2. .148 of consensus"
repeat_region
1136. .1187
/note="26 copies 2 mer gt 98% conserved"
repeat_region
2960. .3272
/note="AluSg repeat: matches 1. .310 of consensus"
repeat_region
4144. .4270
/note="AluX repeat: matches 1. .114 of consensus"
repeat_region
4319. .4494
/note="AluX repeat: matches 118. .293 of consensus"
repeat_region
5662. .6438
/note="L1MEC repeat: matches 272. .1095 of consensus"
repeat_region
6916. .8080
/note="L1MEC repeat: matches 1168. .2367 of consensus"
repeat_region
9934. .10087
/note="MIR repeat: matches 25. .184 of consensus"
repeat_region
10353. .10481
/note="MIR repeat: matches 65. .194 of consensus"
repeat_region
11077. .11372
/note="AluX repeat: matches 1. .294 of consensus"
repeat_region
11600. .11639
/note="20 copies 2 mer tc 95% conserved"
repeat_region
12479. .12608
/note="65 copies 2 mer at 90% conserved"
repeat_region
13433. .13876
/note="MTRC repeat: matches 1. .466 of consensus"
repeat_region
14372. .14576
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region
14577. .14633
/note="AluSg repeat: matches 243. .299 of consensus"
repeat_region
14843. .14943
/note="MIR6 repeat: matches 9. .109 of consensus"
repeat_region
15404. .15613
/note="MIR repeat: matches 3. .213 of consensus"
repeat_region
16626. .16705
/note="MIR repeat: matches 60. .140 of consensus"
repeat_region
16928. .16993
/note="33 copies 2 mer gg 66% conserved"
repeat_region
17476. .17569
/note="MER81 repeat: matches 2. .114 of consensus"
repeat_region
17719. .18069
/note="L1MC4 repeat: matches 7617. .7977 of consensus"
repeat_region
18117. .18345
/note="MIR repeat: matches 7. .262 of consensus"
repeat_region
19169. .19400
/note="MIR repeat: matches 5. .239 of consensus"
repeat_region
19436. .19627
/note="L1MCS repeat: matches 7728. .7917 of consensus"
repeat_region
19642. .19716
/note="MIR repeat: matches 48. .131 of consensus"
repeat_region
19969. .20010
/note="21 copies 2 mer tg 100% conserved"
repeat_region
20623. .20704

/note="41 copies 2 mer gt 85% conserved"
repeat_region
20738. .20773
/note="U2 repeat: matches 1. .36 of consensus"
repeat_region
20802. .20863
/note="MTR1J repeat: matches 1. .62 of consensus"
repeat_region
20889. .21263
/note="MTR1F repeat: matches 188. .541 of consensus"
repeat_region
21463. .21618
/note="MIR repeat: matches 46. .192 of consensus"
repeat_region
22019. .22326
/note="Aluub repeat: matches 1. .306 of consensus"
repeat_region
22381. .22564
/note="U2 repeat: matches 2453. .2629 of consensus"
repeat_region
22896. .23174
/note="MTR1J repeat: matches 117. .413 of consensus"
repeat_region
23215. .23346
/note="MIR repeat: matches 48. .188 of consensus"
repeat_region
23388. .23532
/note="U2 repeat: matches 2097. .2230 of consensus"
repeat_region
23533. .23837
/note="Aluub repeat: matches 1. .303 of consensus"
repeat_region
23838. .24137
/note="U2 repeat: matches 1754. .2097 of consensus"
repeat_region
24291. .24581
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region
24653. .24850
/note="MIR repeat: matches 1. .200 of consensus"
misc_feature
26189. .26685
/note="Cpg island"
/evidence="not_experimental"
29453. .30918
/note="Cpg island"
/evidence="not_experimental"
31084. .31156
/note="MIR repeat: matches 65. .138 of consensus"
repeat_region
31618. .31876
/note="Aluuo repeat: matches 29. .275 of consensus"
repeat_region
32767. .32830
/note="MIR repeat: matches 76. .139 of consensus"
repeat_region
33050. .33178
/note="43 copies 3 mer tcc 72% conserved"
35112. .36201
/note="Cpg island"
/evidence="not_experimental"
36551. .36604
/note="27 copies 2 mer ac 94% conserved"
repeat_region
37588. .38254
/note="L1PA15 repeat: matches 5480. .6157 of consensus"
repeat_region
38265. .38561
/note="AluX repeat: matches 1. .300 of consensus"
repeat_region
39214. .39423
/note="MIR repeat: matches 3. .219 of consensus"
repeat_region
39489. .39801
/note="AluX repeat: matches 1. .292 of consensus"
repeat_region
40050. .40189
/note="MIR repeat: matches 1. .144 of consensus"
repeat_region
41057. .41290
/note="MIR repeat: matches 8. .240 of consensus"
repeat_region
41296. .41460
/note="U2 repeat: matches 1916. .2416 of consensus"
repeat_region
41506. .41944
/note="U2 repeat: matches 1448. .1779 of consensus"
repeat_region
42388. .42698
/note="U2 repeat: matches 1448. .1779 of consensus"
repeat_region
44193. .44579
/note="THE1C repeat: matches 1. .371 of consensus"
repeat_region
44600. .44956
/note="L1R16A repeat: matches 90. .445 of consensus"
repeat_region
45240. .45300
/note="MERS8A repeat: matches 37. .97 of consensus"
repeat_region
45798. .45909
/note="L1RA1 repeat: matches 90. .192 of consensus"
repeat_region
46826. .46871

/note="23 copies 2 mer gt 97% conserved"
repeat_region 50684..50969
/note="AluX repeat: matches 1..287 of consensus"
repeat_region 50980..51291
/note="AluX repeat: matches 1..312 of consensus"
repeat_region 52222..52519
/note="AluX repeat: matches 1..300 of consensus"
repeat_region 54065..54260
/note="L1M4 repeat: matches 3865..4055 of consensus"
repeat_region 54261..54432
/note="FAM repeat: matches 2..167 of consensus"
repeat_region 54433..54629
/note="L1M4 repeat: matches 3652..3865 of consensus"
repeat_region 54648..54862
/note="L1M4 repeat: matches 11..217 of consensus"
repeat_region 54863..55236
/note="L1M4 repeat: matches 1..365 of consensus"
repeat_region 55237..55700

Query Match 35.3%; Score 515.8; DB 9; Length 165110;
Best Local Similarity 66.8%; Pred. No. 2e-105;
Matches 820; Conservative 0; Mismatches 357; Indels 50; Gaps 4;

QY 20 GTCCCTGGGCCCCCGTTGCTGATTGGCCCGTGCACAGGACAGCCCGGACGACGCT 79
DB 30814 GGCCCCAGGCCCCCGGCGCTGATTGGCCCGTGCACAGGACAGCCCGGACGACGCT 30755
QY 80 CCTGATCCGGGAGAGCAGATTAAGGTCGCCAGGGGACACAGATTAGAGCTCAGAGT 139
DB 30754 CTTGGCCCGGGGAGAGCAGATTAAGGTCGCCAGGGGACACAGACTGTGCTCAGAGAA 30695
QY 140 CCTCTGGGCTCACCACATGC-----ACAGAGGCCAGAGACCCCTC 181
DB 30694 TCCCTGGGCTCACCACATGC-----ACAGAGGCCAGAGACCCCTC 30635
QY 182 CGAGCTTTCTTGTCTGCTCCAGACGCAATTACT-CCAGGCGAGGGCCCTGACGCTAG 240
DB 30634 CTCTCTTTCTTTCTCTTTGGGGCTGGGCACTCCAGGCGGGCGCTCAGCTCAG 30575
QY 241 CAAACTTCGAGCAGCAGC-----AGGGGTTACGTAATCCACCGCTGT 285
DB 30574 CTGAACCTTCGAGCAGCAGC-----AGGGGTTACGTAATCCACCGCTGT 30515
QY 286 TGACTCTGACACACCGCAGCTCTGTCTTTTGAAGCCCGAGTAAGTAAGTAATTA 345
DB 30514 TATTCTTTGGCCCGGTAGAAAGTAATTTTGAAGGCTCCGAGGAGCGGCAAGGAA 30455
QY 346 GGAACCTCCAAAGGTTAGAAAGGAGGAGTGGGCGTACT-----CT 389
DB 30454 AGAGGAGATCTCTGACCCAGCGGGGCTGGAGATGCTGTTTTTTTCCACCT 30395
QY 390 AGTCCCGGCTGAGTGAAGCTCTTAAGTCAGAGACTGTACACCCCTTCATTTTTC 449
DB 30394 AGCTTCGAGATCGGAGATGCGCCCAAGTACCGAATCAAACTTACCTTCTGACCC 30335
QY 450 AACCTCAGATGAGCGCTCATCTTGAATGAGCCACCATCAAGTGTCCCAAGAGCC 509
DB 30334 CGCCGTGATTAACGCTTAACCTTCGAGTGGCCCACTGTCAAGTACCCCTGAGAG 30275
QY 510 CAGCAACCTTTCCGGAGCCTCGACCAAGAGTGTCTCAATTCACCCACCT 569
DB 30274 GAGCGGTCTTCCCGAGAGCTCGAAGAGAGTGAAGTGTGCTCCACCGCCGCC 30215
QY 570 AGCCCCACTCTGTATCCAGAGGATGTCTCCAGAGCAAGAGAGTGTGCGGAGAGCA 629
DB 30214 AGCCCCACTCTGTATCCAGAGGAACTGTCCAGAGGAGAGTGTGCGGAGAGCA 30155
QY 630 TCGAGGAAGCTCGGTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 689
DB 30154 CCGAGGAAGCTCGGTGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 30095
QY 690 AGCAAGCAGCAGAGAGCGGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 749
DB 690 AGCAAGCAGCAGAGAGCGGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGC 749

DB 30094 AGCAAGCAGCAGAGAGCGGCGGAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGC 30035
QY 750 AACCTTAATCTCGGCTGATGATGCGTGGCGGCTGTCTGCCACCTTCCGAGATGACGCC 809
DB 30034 AACCTTAATCTCGGCTGATGATGCGTGGCGGCTGTCTGCCACCTTCCGAGATGACGCC 29975
QY 810 AAATTTACAAAGATGAGACCTTGGCTTCCGCCCAACATACATTTGGGCACTGACTGAG 869
DB 29974 AAGCTCACCAAGATGAGACCTTGGCTTCCGCCCAACATACATTTGGGCACTGACTGAG 29915
QY 870 AAGCTGCGCATGAGGAGACCAAGCTTCAAGGCCCCCGGACCCCTGTGGCGGAG 929
DB 29914 AAGCTGCGCATGAGGAGACCAAGCTTCAAGGCCCCCGGACCCCTGTGGCGGAG 29855
QY 930 CTGGGAAGCCCGGAGAGGGGCTCCAGCGGCACTGAGGAGCTTATCTACTCCCAATTTC 989
DB 29854 GAGCTGGGACACCCAGCGGCTTCCCGGAGACTGGGGGTCTCTTACTCCCAATCTCC 29795
QY 990 CAAGCTGTAGCTGAGCCCGACAGCTCATTTGAGAGAGTTCCTGCTGACAGTCCG 1049
DB 29794 CAGGCTGGACCTGAGTCCCGCGCGTCTGTGAGAGGAGCAGCCGGGCTGAGGAGCC 29735
QY 1050 AGCTCCCATCTCTGTCTGCTCCGCGACCTGTGTCTCAAGCTTTGTGAAGGCC 1109
DB 29734 ACCTTTCCGCTCTGTGAGCCAGGCACTGTGCTTCTCAAGTCTTGTGAAGGAC 29675
QY 1110 CAAACAGGCGCTGGGCGGCTGGGCGCTGGCAGAAAGGAGGAGGAGGAGGAGGAGGAG 1169
DB 29674 CTGTCTGTCTGTGGCTGTGAGTGAAGGTAAGGAGGAGGAGGAGGAGGAGGAGGAG 29615
QY 1170 GGAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196
DB 29614 AGAGGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 29588

RESULT 13
AC021954/C
LOCUS
DEFINITION Homo sapiens chromosome 10 clone RP11-57E12 map 10, WORKING DRAFT
ACCESSION AC021954.3
VERSION AC021954.3 GI:7417809
KEYWORDS HTG; HTGS; PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 173341)
Birtten,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 10, clone RP11-57E12
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 173341)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N., Anderson,S., Baldwin,J., Barna,N., Beckert,R., Beda,F., Boguslavsky,I., Boukhalter,B., Brown,A., Burkett,G., Castle,A., Chappel,V., Colangelo,M., Collins,S., Collymore,A., Cooke,P., Dekrellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J., Ferreira,S., Grant,G., Hagos,B., Heald,D., Horton,L., Howland,J., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Landers,T., Lehoczy,J., Levine,R., Liu,G., Locke,K., MacDonald,P., Margulis,N., McEwan,P., McGuck,A., McKernan,K., McNeeters,R., Meldrum,J., Meneus,L., Morrow,J., Naylor,J., Norman,C.H., O'Connor,F., O'Donnell,P., Olivari,T.M., Peterson,K., Piere,N., Pisan,C., Pollara,V., Raymond,C., Riley,R., Rothman,D., Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Testave,S., Theodore,J., Titrill,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 173341)


```

misc_feature 74797..81326
                /note="assembly_fragment"
misc_feature 81427..89446
                /note="assembly_fragment"
misc_feature 89547..96459
                /note="assembly_fragment"
misc_feature 96560..106498
                /note="assembly_fragment"
misc_feature 106599..116986
                /note="assembly_fragment"
misc_feature 117087..128890
                /note="assembly_fragment"
                clone_end:77
                vector_side:right"
misc_feature 128991..147290
                /note="assembly_fragment"
misc_feature 147391..173341
                /note="assembly_fragment"

Query Match      35.1%; Score 512.6; DB 2; Length 173341;
Best Local Similarity 66.7%; Pred. No. 1.1e-104;
Matches 818; Conservative 0; Mismatches 359; Indels 50; Gaps 4;

Qy 20 GTCCCTGGGCCCCCGTGTGATGAGCCCGTGGCAAGGAGGAGCCCGGAGGAGCGCT 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7254 GAGCCCAAGGCGCCGCGCTGATGAGCGGTGGCGGGGAGAGCCCGGAGGAGCGCT 7195
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 80 CCTGTCCGGGAGAGAGATTAAGCTGCCAGGGGAGACACAGATTACAGCTCAAGAGT 139
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7194 CTTGGCCGGGAGAGATTAAGCTGCCAGGGGAGACACAGATTACAGCTCAAGAGT 7135
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 140 CCTCTGGGCTCACCATGCG-----ACAGAGCGGAGGAGCCCGCTC 181
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7134 TCCGCGGCTCTCACCGCGCGCTCGAGAGAGAGCTGACAGAGGCTCTCGAGCCCACTT 7075
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 182 CGAGCTTCTTCTGCTCTCCAGACGCAATTACT-CCAAGCGAGGCGCGCTGACAGCTCAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7074 CTCTCTTCTTCTCTTCTTGGGCGTGGGCAACTCCAGGGGCGCGCTGACAGCTCAG 7015
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 241 CAAAATTCTGAAGCAGCAG-----AGGGGTTCAGCTATCCAGCGCTCT 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 7014 CTGAATTCTGAGCAGCAGAGCGCGCTGAGCTCCACGCGCTCTCTCTCATCTCTCTC 6955
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 286 TGACTCTGACCAAGCGAGCTCTGTTCTTTGAGCCCGAGTAAGTAAGTAACTTTA 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6954 TATTTTTCGCGCGCTGAAGAGTAATTTTTCAGGCGCTTCGAGGAGCGGAGGAGAA 6895
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 346 GGAACCTCCAAAGGAGTAAGAGAGGAGTGGGTGGCGTACT-----CT 389
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6894 AGAGGGATCTCTGACCCAGCGGGGCGTGGAGATGAGTGTGTTTGTTCACACT 6835
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 390 AGTCCCGGTGAGTGAAGCTTAAGTCAAGACTGTCAACCCCGCTTCATTTTTC 449
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6834 AGCTCTGGAATCGCGAGCTGCGCCAGTGAAGAGTCAAACTTACCTTCCCTGAGCC 6775
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 450 AACCTCAGAGTGGGCTCATCTTCATGATGGCCACCATCAGTGTCCCAAGAGAC 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6774 CGCCGTAGATGACCTCAACCTCGGGTGGCCACGTCACAGTACCCGTAGACG 6715
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 510 CAGCAACCTTTCCCGAGCTCGAGACACAGAAAGTCTCAAGTTCCAAATTCACCCCACT 569
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6714 GAGCGGTCTTCCCGAGAGCTCGAAGAGAGTGAAGTCTGCCACAGTCCGCGCCCGCC 6655
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 570 AGCCCACTCTCTGATCCGAGGAGTGTCTCCAGAGCAAGAGAGTGAAGTCCGAGGAGCA 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6654 AGCCCACTCTGATCCGAGGAGTGTCTCCAGAGCAAGAGAGTGAAGTCCGAGGAGCA 6595
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 630 TCGAGAGAGTCCGTGCGGGGCGGAGGCGGCAACAGGCCCAAGAGGAGTGGCACTG 689
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6594 CCGAGAGAGTCCGTGCGGGGCGGAGGCGGCAACAGGCCCAAGAGGAGTGGCACTG 6535
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 690 AGCAAGCAGCAGAGCGGCGCAAGAGCGCAAGAGCGGAGCGCAACCGCATGAC 749
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db 6534 AGCAAGCAGCAGAGCGGAGTCCGCGCAAGAGAGCCCAAGCAGCCGCAATCGAATGCAC 6475
Qy 750 AACCTTAATCCGGCTGATGAGTGGCGGTGTCTGTCACCTTCCGAGTGAAGGCC 809
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6474 AACCTTAATCCGGCTGATGAGTGGCGGTGTCTGTCACCTTCCGAGTGAAGGCC 6415
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 810 AAACCTTAAGAGATGAGAGCTTGGGCTTGGCCCAACTAATTTGGGCACTGACTCAG 869
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6414 AAGCTCACAAAGATCGAGAGCTGCTGCGCCCAACTAATCTGGGCGTGAATCAA 6355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 870 ACGTGGCATGAGGAGCACAAGTCTTACAGGCCCCCAAGCCCCCTGTGCTGTGGAG 929
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6354 ACGTGGCATGAGGAGCACAAGTCTTACAGGCCCCCAAGCCCCCTGTGCTGTGGAG 6295
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 930 CTGGGAAGCCCGGAGGAGGCTCCAGGCGGAGCTGAGGCTCTATCTACTCCCAAGTTCC 989
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6294 GAGCTGGGAGCCAGAGCGGCTTCCCGGGAGCTGGGGGTCTCTACTCCCAAGTTCC 6235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 990 CAAGCTGTAGCTTGAAGCCCAAGCTTCAATTTGAGAGAGTTCCCTGCTGAGAGTCCC 1049
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6234 CAGCTGTGAGCTGAGTCCCGCGCGCTGCTGAGAGAGCAGCCGCGCTGTGGAGGCC 6175
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1050 AGCTCCCATCTCTGTCTTCCGGGAGACCTGTGTCTTCAAGCTTCTTGTGAAGGCC 1109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6174 ACCTTTCCGCTCTTGAAGCCAGGAGTGTGCTTTCTCAATTTCTGTGAAGGAC 6115
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1110 CAAACAGGCCCCGTGGGCGGTGGGCGCTGGCAGAAAGGAGGAGTCAAGCTGTCTGAAT 1169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6114 CTGTCTGCTGCTGGCTGTGGGTCTTAAGGTAAAGAGAGAGAGAGAGAGAGAGAGCCGT 6055
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 1170 GGAAGTATGAGAGCACTGAGCATC 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 6054 AGAGGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 6028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
HSA133776      1330 bp      DNA      linear      PRI 19-JUN-1999
LOCUS          Homo sapiens gene for neurogenin 3.
DEFINITION
ACCESSION      AJ133776
VERSION        AJ133776.1 GI:5123782
KEYWORDS       bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE         Homo sapiens (human)
ORGANISM       Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE      1
AUTHORS        Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
                Maillet,J.
TITLE          The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
                expression pattern is identical to that of its murine counterparts
                Unpublished
                2 (bases 1 to 1330)
REFERENCE      2
AUTHORS        Ravassard,P.
TITLE          Direct Submission
                Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMRC 9923, Hopital
                de la Pitie Salpetriere, Bat. CERVI, 83 Bd. de l'Hopital, 75013
                PARIS, FRANCE
FEATURES
source
    1..1330
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
    1..1330
        /gene="ngn3"
        join(<1..157,321..>1330)
    1..157
        /gene="ngn3"
    1..157
        /number=1
    158..320
        /gene="ngn3"
        /number=1
    intron

```


THIS PAGE BLANK (USPTO)


```

FT      /product= "Ngn3"
FT      /note= "Neurogenin 3"
PN      WO200059936-A1.
XX      12-OCT-2000.
XX      28-MAR-2000; 2000WO-US08436.
XX      06-APR-1999; 99US-0128180.
XX      (REGC ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI: 2000-664989/64.
XX      P-PSDB; AAY85618.
PT      Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT      them, useful for diagnosis, prevention and treatment of diabetes
PT      mellitus and to identify individuals at risk of diabetes -
XX      Claim 18; Page 49-50; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAc61089 encodes the Ngn3
CC      protein AAY85617. The Ngn3 gene is located at chromosome position
CC      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC      protein sequences, and includes an antibody recognising the Ngn3 protein.
CC      Also included in the invention is a method for identifying an islet cell
CC      precursor, the method involves analysing a cell for the expression of the
CC      Ngn3 gene product, where detection of the product is indicative of an
CC      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC      reagent for detecting (in a subject) a predisposition to a defect in
CC      pancreatic islet cell function or formation associated with a defect in
CC      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC      precursor cells expressing Ngn3, and to alter cellular differentiation in
CC      culture in vivo to produce new beta-cells to treat patients with diabetes
CC      mellitus. The present sequence represents the murine Ngn3 genomic DNA
CC      sequence.
XX      Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;
SQ
Query Match      61.7%; Score 900.6; DB 21; Length 1861;
Best Local Similarity 86.2%; Pred. No. 4,9e-235;
Matches 1094; Conservative 0; Mismatches 129; Indels 46; Gaps 7;
QY      4 GGTAGCGAGAGAGAGAGTCCCTGGGCCCCCGTGTGATTTGGCCCGTGACACAGGACGA 63
DB      594 GGGCTCGGTGGCGGAGCCCTCGGCCCCCTCCGCTGATTGGCCCGTGTGACAGGACGA 653
QY      64 GCCCGGAGGACGCTCTGTGTC -GGGCAAGACAGATTAAGCCGACAGGAGACACAG 122
DB      654 GCCCGGAGGACGCTCTGTGTCGGGGGAGAGAGATTAAGCCGACAGGAGGACACAG 713
QY      123 ATT-----AGAGCTCAGAAAGTCCCTCTGGGTCTACACACTGCA -CAGAGGCCAGAGACC 177
DB      714 ACTTGACATGCACTCAGAAATCCCTCTGGGTCTATCACTGACAGAGAGTGTCAGTACTT 773
QY      178 CCTCGAGCTTCTTGTGCTGCTTCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACACT 237
DB      774 CCTGGAGCTTTTCTACGACTTCAGACGCAATTTACTCCAGGCGAGGGCGCTTGACACTT 833
QY      238 CAGCAAACTTTCAGAGCGAGAGAGGGTTACAGCTATCCACGCTGCTTTGACTGTGACC- 296
DB      834 TAGCAGAACTTCAGAGGAGAGAGAGGCTCAGCTATCCACGCTGCTTTGACACTGACACC 893
QY      297 -----ACCGGAGCTCTGTGTTCTTTTGAAGCCGAGATA 330
DB      894 TATCAGCTGCTGCTGTCTACTGACTGACTGCTGCTCTCTATTTCTTTTGAAGTCGAGAG-A 952
QY      331 ACTAGTAACATTTAGAACTCTCAAGGGTAGAAGAGGAGTGGGT-----GG 380
DB      953 ACTAGTAACATTTAGAACTCTCAAGGGTAGAAGAGGAGGCGCGGGGTGTGTGGG 1012

```

```

QY      381 GCGTACTTGTAGTCCCGGAGTGAAGTACCTTAAGTCAGAGACTG--TCACACCCCCCTTC 438
DB      1013 GGATACTCTGGTCCCGGTGAGTGAACCTTAAGTCAGAGGCTGGACACACACACCTTC 1072
QY      439 CATTTTTCCTCACTCAGATGAGTGGCGCTTATCCCTTGGATGCGCCACCATCCAAAGTGT 498
DB      1073 CATTTTTCCTCACTCAGATGAGTGGCGCTTATCCCTTGGATGCGCCACCATCCAAAGTGT 1132
QY      499 CCCAAGAGACCCACCAACCCCTTCCGGAGGCTGCGACCAAGAGTGTGCTCAATT 558
DB      1133 CCCAAGAGACCAACCAACCTTTCGGAGCTTGGACCAAGAGTGTGCTCAATT 1192
QY      559 CCACCCACCTAGCCCACTCTCGTACCGAGGAGCTGCTCGAAGACAGAAACAGTGAAT 618
DB      1193 CCACCCACCTAGCCCACTCTCATCTAGGAGCTGCTCGAAGACAGAAAGTGTGACT 1252
QY      619 GCCGAGGACATCGAGAGAGTCCGTGCGGGCGCGGAGGGCGCAACAGCCCAAGAGCG 678
DB      1253 GCCGAGGAGCTCGAGGAGAGTCCGCGCCGACCGGAGGGCGCAAGAGGCCCAAGAGCG 1312
QY      679 AGTTGGACATGAGAACGACGAGAGCGAGCGGCGCAAGAGGCCAAGACCGGAGAGCGCA 738
DB      1313 AGTTGGACATGAGAACGAGAGAGCGGCGCAAGAGGCCAAGATGATGAGAGCGCA 1372
QY      739 ACCGATGACACAACTTAACTCCGCGTGAATGCGTGGCGGTGCTCCTGCGCCACTTCC 798
DB      1373 ATCGATGACACAACTTAACTCCGCGTGAATGCGTGGCGGTGCTCCTGCGCCACTTCC 1432
QY      799 CGATGACGCGCAACTTAAAGATGAGAACCTCGCGCTTGGCCACCAACTCAATTGGG 858
DB      1433 CGATGACGCGCAACTTAAAGATGAGAACCTCGCGCTTGGCCACCAACTCAATTGGG 1492
QY      859 CACTGACTCAGACGCTGCGATGACGAGACCAAGCTTTCAGGCGCCGAGGCCCTGTGC 918
DB      1493 CACTGACTCAGACGCTGCGATGACGAGACCAAGCTTTCAGGCGCCGAGGCCCTGTGC 1552
QY      919 CCTGTGGGAGTGGAGAGCCCGGAGGGGCTCCAGCGGCGCATGAGGGCTCTAATCTACT 978
DB      1553 CCTGTGGAGAGTGGAGAGCCCGGAGGGGCTCCAGCGGCGCATGAGGGGCTCTAATCTACT 1612
QY      979 CCCGAGTTTCCAGAGCTGTAGCTTGAAGCCGACAGCTCTATTGAGAGATTTCCCTGGCC 1038
DB      1613 CCCGAGTTTCCAGAGCTGTAGCTTGAAGCCGACAGCTCTATTGAGAGATTTCCCTGGCC 1672
QY      1039 TGCAGGTGCCAGTCTCCCATCTGTCTGTCTCCGGGACACCTTGCTGTTCACACTTCT 1098
DB      1673 TGCAGGTGCCAGTCTCCCATCTGTCTGTCTCCGGGACACCTTGCTGTTCACACTTCT 1732
QY      1099 TGTGAAGGCGCCAAACAGAGCCCTGGGCGGTGGCGCTGGCAGAAAGGAGGAGTCAAG 1158
DB      1733 TGTGAAGAGACTGTCTGTGCTGTGGTGTGGTGTGCTTGAAGAGGAGGAGACAGAG 1792
QY      1159 CTGTCTGAATGAGAGGTAGTGAAGGACCTGAGCATCTGCGCCCTTCTGGCTTTCAATTA 1218
DB      1793 CCGTCTGAGAGTGGAGGTAGTGAAGGCTTCAACATCTGCGCTTCTGCTTCACTA 1852
QY      1219 GTCAAGTTC 1227
DB      1853 CTTGAGATCC 1861

```

RESULT 3
 AAD46872
 ID AAD46872 standard; DNA; 1860 BP.
 AAD46872;
 AC AAD46872;
 XX
 DT 27-JAN-2003 (first entry)
 XX
 DE Murine neurogenin 3 (Ngn3) gene.
 XX
 KW Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;

KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
 KW islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
 OS Mus musculus.
 XX Key Location/Qualifiers
 XX FT CDS 1093..1737
 XX FT /*tag= a
 XX FT /product= "Murine Ngn3 protein"
 XX PN WO200274045-A2.
 XX PD 26-SEP-2002.
 XX PF 20-MAR-2002; 2002WO-US11166.
 XX PR 20-MAR-2001; 2001US-0817360.
 XX PA (REGC) UNIV CALIFORNIA.
 XX PI German MS, Lin J;
 XX DR WPI; 2002-759853/82.
 XX DR P-PSDB; AAE29278.
 XX PT Producing a mammalian islet cell for treating diabetes mellitus
 XX PT comprises introducing into a mammalian cell a nucleic acid molecule
 XX PT encoding neuroendocrine basic helix-loop-helix transcription factor -
 XX PS Example 3; Page 89-90; 108bp; English.
 XX CC The invention relates to a method for producing a mammalian islet cell.
 XX CC The method comprising introducing into a mammalian cell a nucleic acid
 XX CC molecule encoding an islet transcription factor for expression of the acid
 XX CC islet transcription factor in the cell and for production of islet cell
 XX CC phenotype in the cell. The islet transcription factor is a neuroendocrine
 XX CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
 XX CC for treating type 2 diabetes mellitus and for replacing beta cells lost
 XX CC to autoimmune destruction in individuals with type 1 diabetes. The method
 XX CC is useful in cell therapy. The present sequence is murine neurogenin 3
 XX CC (Ngn3) gene.
 XX CC
 XX SQ Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;
 Query Match 61.6%; Score 899.6; DB 24; Length 1860;
 Best Local Similarity 86.2%; Pred. No. 9.2e-235;
 Matches 1093; Conservative 0; Mismatches 129; Indels 46; Gaps 7;

Db 953 ACTAGTAACTTTCGAAATCCAAAGGATGAGGGGCGCGGGTGTGTGGG 1012
 QY 381 GCGTACTTACTCCCGGTGAGTGAACCTCTAAGTCAGAGCTG--TCACACCCCTTC 438
 Db 1013 GGAATCTGTGTCCCGGTGAGTGAACCTCTAAGTCAGAGCTGACACACACACCTTC 1072
 QY 439 CATTTTTCCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 498
 Db 1073 CATTTTTCCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 1132
 QY 499 CCAAGAGACCCAGCAACCTTTTCCGAGGCTTGGACCAAGAGTGTCAATT 558
 Db 1133 CCCCAGAGACACAAACACTTTTCCGAGGCTTGGACCAAGAGTGTCAATT 1192
 QY 559 CCACCCCACTTACCCCACTTCTGTAACCAAGAGTGTCAATT 618
 Db 1193 CCACCCCACTTACCCCACTTCTGTAACCAAGAGTGTCAATT 1252
 QY 619 GCCGAGGACATCGAGAGTCCGTGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGG 678
 Db 1253 GCCGAGGACCTCGAGAGTCCGTGCGCGCGCGAGGAGGAGGAGGAGGAGGAGGAGG 1312
 QY 679 AGTTGGCACTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 738
 Db 1313 AGTTGGCACTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1372
 QY 739 ACCGATGACAACTTACCTCCGCGGTGAGTGGCGGTGCTGCGCACTTC 798
 Db 1373 ATCGATGACAACTTACCTCCGCGGTGAGTGGCGGTGCTGCGCACTTC 1432
 QY 799 CGGATGAGCCAACTTACCAAGATCGAGACCTTCCGCTTCCGCACTTATTTGGG 858
 Db 1433 CGGATGAGCCAACTTACCAAGATCGAGACCTTCCGCTTCCGCACTTATTTGGG 1492
 QY 859 CACTGACTCAAGCTGCGCATAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 918
 Db 1493 CACTGACTCAAGCTGCGCATAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1552
 QY 919 CCTGTGGAGTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 978
 Db 1553 CCTGTGGAGTCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1612
 QY 979 CCCCAGTTTCCCAAGCTGTAGCTGAGCCCAAGCCTCAATTGAGAGATTCCTGAGC 1038
 Db 1613 CCCCAGTTTCCCAAGCTGTAGCTGAGCCCAAGCCTCAATTGAGAGATTCCTGAGC 1672
 QY 1039 TGCAGTCCCAAGCTCCCATCTGTCTGCTCCGCGGACCTGTGTCTCAACTTCT 1098
 Db 1673 TGCAGTCCCAAGCTCCCATCTGTCTGCTCCGCGGAGCACTGTGTCTCAACTTCT 1732
 QY 1099 TGTGAAGGCCCCAAACAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1158
 Db 1733 TGTGAAGGAGCCTGT 1792
 QY 1159 CTGTCTGAATGAAAGT 1218
 Db 1793 CGGTCTGAGTGGAGGAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1852
 QY 1219 GTCAAGTCT 1226
 Db 1853 CTTGATTC 1860

RESULT 4
 AAF27254
 ID AAF27254 standard; cDNA; 5567 BP.
 XX AAF27254;
 AC AAF27254;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.

XX	Atonal; homologue; orthologue; atonal-associated protein; deafness;
KM	hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW	cellular proliferation; cerebellar granule neuron; gene therapy;
KW	mechanoreceptive cell growth; auditory; osteopathic; cyostatic;
KX	transgenic animal; sa.
OS	Mus musculus.
XX	WO20073764-A2.
EN	07-DEC-2000.
PD	01-JUN-2000; 2000MO-US15410.
PE	01-JUN-2000; 99US-0137060.
XX	PR 19-JAN-2000; 2000US-0176993.
XX	(BAYU) BAYLOR COLLEGE MEDICINE.
PA	Zoghbi HY, Belien H, Birmingham N, Hassan B, Ben-Arie N,
PI	WPI; 2001-032190/04.
XX	DR P-PsDB; AAB60350.
XX	The invention relates to the use of atonal-associated nucleic acid or
PT	any of its homologs or orthologs, for the treatment of e.g. deafness,
PT	osteoarthritis and abnormal cell proliferation -
XX	Disclosure; Page -, 142pp; English.
PS	The invention relates to the use of atonal-associated nucleic acid or
XX	amino acid sequence, or any of its homologues or orthologues as
CC	therapeutic agents for the treatment of deafness, partial hearing loss,
CC	vestibular effects due to damage or loss of inner hair cells.
CC	osteoarthritis and abnormal cell proliferation. The invention also
CC	encompasses methods of screening for compounds which affect the
CC	expression of an atonal-associated nucleic acid sequence in an animal,
CC	and a transgenic animal in which an allele of a native atonal-associated
CC	gene is replaced by a heterologous nucleic acid sequence, thus
CC	inactivating the atonal-associated allele. The nucleic acids or proteins
CC	may be used in a method of treating an animal for hearing impairment,
CC	joint disease, balance disorders, abnormal cell proliferation, or other
CC	disease related to loss of a functional atonal-associated nucleic acid or
CC	protein. They may particularly be used to treat an animal with a
CC	deficiency in cerebellar granule neurons or their precursors, and may
CC	also be used in promoting mechanoreceptive cell growth and generating
CC	hair cells. The present sequence represents an atonal-associated nucleic
CC	acid sequence referred to in the invention.
CC	Note: The present sequence is not shown in the specification, but
CC	was obtained from GenBank.
XX	Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other:
SQ	
Query Match	56.2%; Score 819.8; DB 22; Length 5567;
Best Local Similarity	87.1%; Pred. No. 7.5e-213;
Matches 997; Conservative	0; Mismatches 102; Indels 46; Gaps 7;
OY	4 GGTACGAGAGAGAGATGCCCTGGGGCCCCCGTTCGTAATTGGCCCGTGACACAGCACA 63
Db	4424 GGCGTCGGTGGCGAGAGCCCCCGGGCCCCCTCGCTGAATGGCCCCGTGGTGAGGCACGA 4483
OY	64 GCCCGGAGAGCAAGCTCCTGATC-CGGGCGAGAGAGATTAAGCTGCCAGGGGACACAGC 122
Db	4484 GCCCGGAGAGCAAGCTCCTGATC-CGGGCGAGAGAGATTAAGCTGCCAGGGGACACAGC 4544
OY	123 ATT---AGCAGCTCAAAATGCTCTTGGGTTCACCACTGCA-CAGAAGCCGAGAACCC 177
Db	4544 ACTTCATGCAAGCTCAAAATCCCTCTGGGTTCATCACTGCAGCAGTAGTGAGTAAGTACT 4603
OY	178 CCTCGAGGCTTTGTCGTCGCCCTCCAGAGCAATTTAATCCAGGAGAGGGCCGCTCAGCT 237
Db	4604 CCTCGAGGCTTTTCTACGACTTCCAGAGCAATTTAATCTCCAGGAGAGGGCCGCTCAGTT 4665

QY	238	CAGCAAAACCTTGAGAGGAGAGAGGGGTTGAGTATCAACGGCTGTGACTGACC-	296
Db	4664	TAGCGAACTTCAGAGGAGAGAGAGGCTAGCTATTCATCTGCTGTGCACTGACC	4723
QY	297	-----ACCCGACCTCTGTGTTTGAAGCCCGAGTA	330
Db	4724	TATCCACTGCTGCTTGTCACTGACGACCTGCTGTCTTATTTCTTTTGAATCGGGAG-A	4782
QY	331	ACTAGGTAACTTTAGGAACCTCCAAAGGTTAGAAAGAGGGAGTG------GGTGG	380
Db	4783	ACTAGGTAACTTCGAAACTCCAAAGGTTAGATGAGGGGCGCGCGGGATGTGTGTGGG	4842
QY	381	GCGAACTTAATCCCGGTGAGTGAGCTCTAAGTACGAGAGACTGTCAACCC--CCTTC	438
Db	4843	GGATATCTGTGTCCTCCGTGCAATGACCTCTTAAGTCAGAGGCTGGCACACACACCTTC	4902
QY	439	CATTTTTCCCAACTCAGATGAGCGCTCACTCCCTTGATGCGCCACCATCAAGTGT	498
Db	4903	CATTTTTCCCAACCGCAGATGAGGCGCTCATCCCTTGATGCGCGTCACTCAAGTGT	4962
QY	499	CCCAAGAGACCGAGAACCTTTTCCCGAGCTTGGACACAGAACTGTCAATTCAATT	558
Db	4963	CCCAAGAGACAACAACCTTTTCCCGAGCTTGGACACAGAACTGTCAATTCAATT	5022
QY	559	CCACCCCACTAGGCCCCCTCTCGATCGAGAGGACTGTCTCCGAGCAGAACAGTACT	618
Db	5023	CCACCCCACTAGGCCCCCTCTCTAATCTTAAGGACTGTCTCCGAGCAGAAATGGTACT	5082
QY	619	GCCGAGGACATCGAGAAAGCTCCGTGCGCGCGAGGCGGCAACAGGCCAAGACG	678
Db	5083	GCCGAGGAGCTTCGAGGAAAGCTCCGCGCGCGAGGCGGCAACAGGCCAAGACG	5142
QY	679	AGTTGGCACTGAGCAAGAGCGGAGAACCGCGCCAGAAAGCCACACGACGGAGCGCA	738
Db	5143	AGTTGGCACTCAGCAAAACAGCGAAAGACCGCGCCAGAAAGCCAAATGATGGAAGCGCA	5202
QY	739	ACCGCATSCAAACCTTAACTCCGCGCTGATGCGTGCCTGCGGTCTCTGCCACCTTCC	798
Db	5203	ATCGCATSCAAACCTCAACTCCGCGCTGATGCGTGCCTGCGGTCTCTGCCACCTTCC	5262
QY	799	CGGATGACGCGAAACTTACAAGATCGAGACCTTCGCTTGCCCACTACATTGTTGGG	858
Db	5263	CGGATGACGCGAAACTTACAAGATCGAGACCTTCGCTTGCCCACTACATTGTTGGG	5322
QY	859	CACGTACCTCAACGCTGGGCACTAGCGGACCAACCTTCTAAGGCCCGCAGGCCCTGTGC	918
Db	5323	CACGTACCTCAACGCTGGGCACTAGCGGACCAACCTTCTAAGGCCCGCAGGCCCTGTGC	5382
QY	919	CCTGTGGGAGCTGGGAGAGCCCGGAGGGGAGCTCAAGCGCGGCACTGGGGCTCTAATCT	978
Db	5383	CCTGTGGGAGAGCTGGGAGAGCCCGGAGGGTGGCTCCAACGGGAGCTGGGGCTCTAATCT	5442
QY	979	CCCCAGTTTCCCAAGCTGTAGCCTGAGCCGCCACAGCCTCAATTGAGAGATTTCCCTGCGC	1038
Db	5443	CCCCAGTCTCCCAAGCGGGTAACTGAGCCGCCACAGCCTCAATTGAGAGATTTCCCTGCGC	5502
QY	1039	TGCAGGTGCCAGCTCCCGCATCTGTCTGTCTCCGGGGACCCCTGTGTCTCAACTCT	1098
Db	5503	TGCAGGTGCCAGCTCCCGCATCTGTGTCTGTCTCCGGGAGCACTGGTGTCTCAAGATCTCT	5562
QY	1099	TGTGA 1103	
Db	5563	TGTGA 5567	
RESULT 5			
AAf27266			
ID AAf27266 standard; cDNA; 861 BP.			
XX AAf27266;			
AC			
DT 24-APR-2001 (first entry)			

[illegible]

OY		660	CGCAACAGAGCCCAAGAGGATGGCACACTGAGCAAGCAGACGACGAAACCAGGCGCAAGAAG	719
Dd		361	CGCAACAGAGCCCAAGAGCGAAGTTGGCATCTCACGAACAACGACGAAGAAGCCGGCGCAAGAAG	420
OY		720	GCCAACGACCGGAGAGCGCAACCGCATGCAACAACCTTAACTCCGCGCTGGATCGCTGC	779
Dd		421	GCCAATGATCGGAGAGCGCAATCGCATGCAACAACCTCAACTCGGCGCTGGATCGCTGC	480
OY		780	GGTGTCTCTGCCCACTTTCCTCGGATGACGCGCAAACCTTAACAAAGATTGAGACCTTGCGCTTC	839
Dd		481	GGTGTCTCTGCCCACTTTCCTCGGATGACGCGCAAACCTTAACAAAGATTGAGACCTTGCGCTTC	540
OY		840	GCCCAACAATAATTGGGGACACTGACTAGACGCTGGCGCATGAGCGAACCAAGACTTCTAC	899
Dd		541	GCCCAACAATAATTGGGGACACTGACTAGACGCTGGCGCATGAGCGAACCAAGACTTCTAT	600
OY		900	GAGCCCGAGCCCCCTGTGCTCCCTGTGGGAGAGCTGGAGAGCCCAGGAGGGGGCTTCAAGCGGC	959
Dd		601	GAGCCCGAGCCCCCTGTGCTCCCTGTGTGGAGAGCTGGGAGAGCCCAGGAGGTGGCTTCAAGCGGC	660
OY		960	GACTGGGGCTCTATCTACTCTCCCACTTTCCTCAAGCTGTGAGCTGAGCTTCAAGCTTCA	101
Dd		661	GACTGGGGCTCTATCTACTCTCCCACTTTCCTCAAGCGGGTAACTGAGCCCCCAAGGCTTCA	720
OY		1020	TTGAGAGAGTTCCTCGGCTGAGGTGCGCCAGCTCCCATCTGTCTGCTCCCGGCGACC	107
Dd		721	TTGAGAGAGTTCCTCGGCTGAGGTGCGCCAGCTCCCATCTGTCTGCTCCCGGCGAGCA	780
OY		1080	CTGTGTCTCTCAGACTTCTTGTGTGAAGGGGCCCAACAGGCGCTGGGCGGTGGCGCTGGCA	113
Dd		781	CTGTGTCTCTCAGACTTCTTGTGTGAAGAACCTGTGTGGCTGTGGGTGGTGGTCTAGTG	840
OY		1140	GAAAGGAGGGAGTCAAGC 1159	
Dd		841	GAAAGGAGGGAGTCAAGC 860	
<hr/>				
RESULT 7				
ID	AAV27050	standard; cDNA; 804 BP.		
XX	AAV27050;			
AC	AAV27050;			
XX				
DT	17-SEP-1998	(first entry)		
XX				
DE	Mouse neurogenin 3 gene.			
XX				
KM	ds; Mouse; neurogenin; expression vector; recombinant protein;			
XX	antibody; neurogenesis.			
OS	Mus sp.			
XX				
FH	Key	Location/Qualifiers		
FT	CDS	160..804		
FT		/tag= a		
FT		/product= "Mouse neurogenin 3"		
XX				
PN	W09813491-A2.			
XX				
PD	02-APR-1998.			
XX				
PF	24-SEP-1997;	97WO-US17048.		
XX				
PR	17-SEP-1997;	97US--0932411.		
PR	27-SEP-1996;	96US--0722570.		
PR	12-NOV-1996;	96US--0030864.		
PR	19-DEC-1996;	96US--0772009.		
XX				
PA	(CALY) CALIFORNIA INST OF TECHNOLOGY.			
XX				
PI	Anderson DJ, Ma Q, Sommer L;			
XX				
DR	WPI; 1998-230702/20.			

DR P-PSDB; AAW54947.
 XX Mouse neurogenesis, useful in neurogenesis - and recombinant nucleic
 PT acids and proteins derived from rat and Xenopus
 XX
 XX Disclosure; Fig 9; 106pp; English.
 XX
 CC The Mouse neurogenin 3 is one of several neurogenin proteins discussed
 CC in the present invention. The neurogenin nucleic acids can be expressed
 CC in a host cell, transformed using an expression vector, to produce
 CC recombinant proteins. The proteins and the antibodies raised against
 CC the proteins are useful in the study of neurogenesis.
 XX
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Query Match 44.7%; Score 652; DB 19; Length 804;
 Best Local Similarity 90.9%; Pred. No. 1.9e-167;
 Matches 731; Conservative 0; Mismatches 60; Indels 13; Gaps 3;

312 TTCTTTGAGCCCGAGTAAGTAACATTAGGAACCTCCAAAGGTAGAGAGGG 371
 2 TTCTTTGAGTGGAG-ACCTAGTAACATTCCGAACTCCAAAGGTAGAGGG 60
 372 AGTGGT-----GGGCGTACTAGTCCCGGTGAGTGAACCTTAAGTCAAGA 421
 61 CGCGGGGGGTGTGTGGGGGATGACTGTGTCCTCCCGGTGAGTGAACCTTAAGTCAAGG 120
 422 CTG--TCACACCCCTCTTCATTTTTCACCACTTCAGATGCGGCTTATCCCTTGAT 479
 121 CTGGGACACACACACTTCCATTTTTCACCACTTCAGATGCGGCTTATCCCTTGAT 180
 480 GCGGCCACCATCAAGTGTCCCAAGAGACCCAGAACCTTCCCGAGACTCCGAGACAC 539
 181 GCGCTCACCATTCAAGTGTCCCAAGAGACCCAGAACCTTCCCGAGACTCCGAGACAC 240
 540 GAAGTGTCTAGTTCATTTCAACCCCACTAGCCCTCTGTCGAGGAGACTGCTCC 599
 241 GAAGTGTCTAGTTCATTTCAACCCCACTAGCCCTCTGTCGAGGAGACTGCTCC 300
 600 GAAGCAGAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGCGAGAG 659
 301 GAAGCAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGCGAGAG 360
 660 GCGACAGAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGAGAG 719
 361 GCGACAGAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGAGAG 420
 720 GCGACAGAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGAGAG 779
 421 GCGACAGAGAGAGTGTCTCCGAGAGATCGAGAGAACTCCGTCGCGCGAGAG 480
 780 GGTGTCTCTGAGCCCTCTCTCCGAGATGCGCAAACTTAAGTAAGTCAAGCCTTCGCTTC 839
 481 GGTGTCTCTGAGCCCTCTCTCCGAGATGCGCAAACTTAAGTAAGTCAAGCCTTCGCTTC 540
 840 GCGCAACTATATTTGGGAGTGAATCGAGAGTGTCTCCGAGAGATCGAGAGAACTTCAC 899
 541 GCGCAACTATATTTGGGAGTGAATCGAGAGTGTCTCCGAGAGATCGAGAGAACTTCAC 600
 900 GCGCCGAGAGAGAGTGTCTCTCTGAGAGTGTGAGAGAGCCGAGAGAGTGTGCTCAAGCGG 959
 601 GCGCCGAGAGAGAGTGTCTCTCTGAGAGTGTGAGAGAGCCGAGAGAGTGTGCTCAAGCGG 660
 960 GACTGGGGCTATCTACTCTCCAGTTCCTCAAGCTGTGAGCTTGAAGCCCAAGCTCA 1019
 661 GACTGGGGCTATCTACTCTCCAGTTCCTCAAGCTGTGAGCTTGAAGCCCAAGCTCA 720
 1020 TTGAGAGATTCCTCTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 1079
 721 TTGAGAGATTCCTCTGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAGTGTGAGAGAG 780
 1080 CTGCTGTCTCAAGACTTCTTGGA 1103

DB 781 CTGCTGTCTCAAGACTTCTTGGA 804

RESULT 8
 AAZ51981
 ID AAZ51981 standard; DNA; 804 BP.
 XX
 XX AAZ51981;
 AC
 DT 04-JUL-2000 (first entry)
 XX
 DE Murine neurogenin-3 (NGN3) nucleic acid sequence.
 XX
 KW Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;
 KW Phox2a protein; neuronal subtype-specific marker; growth factor;
 KW neural differentiation; transplantation; neuronal dysfunction;
 KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;
 KW neuroprotective; nocitropic; anticonvulsant; antiParkinsonian; vulnerary;
 KW cerebrioprotective; immunosuppressant; antineoplastic; ds.
 XX
 OS Mus sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 160..804
 FT /tag=a
 FT /product="Murine neurogenin-3 protein"
 XX
 PN MO200009676-A2.
 XX
 PD 24-FEB-2000.
 XX
 PF 13-AUG-1999; 99MO-US18525.
 XX
 PR 14-AUG-1998; 98US-0096630.
 XX
 PA (CALY) CALIFORNIA INST OF TECHNOLOGY.
 XX
 PI Anderson DJ, Lo L;
 XX
 DR MPI. 2000-256250/22.
 DR P-PSDB; AAY70570.
 XX
 PT Inducing non-neuronal cells to differentiate into neurons and for
 PT non-neuronal cells to express a neuronal subtype-specific marker,
 PT comprising contacting the non-neuronal cells with a vector containing
 PT neurogenin nucleic acid -
 XX
 PS Claim 1; Fig 1J; 76pp; English.
 XX
 CC The patent discloses a method for inducing non-neuronal cells (NNC) to
 CC differentiate into neurons and for NNCs to express a neuronal subtype
 CC -specific marker. Transformed host cells are used as sources of neuronal
 CC and other growth factors; in culture for screening compounds that
 CC modulate neural differentiation or as sources of recombinantly produced
 CC neurogenins and Phox2a proteins for use in transplantation. The cells
 CC also have a variety of in vivo uses, e.g., for transplantation at sites of
 CC neuronal dysfunction e.g. patients with hearing or vision loss due to
 CC optical or auditory nerve damage, brain or spinal cord injuries, and
 CC neurodegenerative disorders e.g. Alzheimer's disease. The present
 CC sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.
 CC NNCs differentiate into neurons through the recombinant expression of a
 CC transcription factor that induces a core program of neurogenesis. Forced
 CC expression of murine NGN3 can elicit expression of at least some neuronal
 CC phenotypic markers even in NNCs.
 XX
 SQ Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

Query Match 44.7%; Score 652; DB 21; Length 804;
 Best Local Similarity 90.9%; Pred. No. 1.9e-167;
 Matches 731; Conservative 0; Mismatches 60; Indels 13; Gaps 3;

312 TTCTTTGAGCCCGAGTAAGTAACATTAGGAACCTCCAAAGGTAGAGAGGG 371

```

Db      2 TTCTTTGATGCGGAG-AACTAGGTAAATTCGAAAATCTCCAAAGGGTGATGAGGGG 60
Qy      372 AGTGGT-----GGCGTACTATAGTCCCGCTGAGTGAAGTCTTAAGTCAGAGA 421
Db      61 GCGCGGGGGTGTGTGTGGGGATTAATCTGTGTCCCGCTGCACTGACTTAAGTCAGAG 120
Qy      422 CTG--TCACACCCCTTTCATTTTTCCTCAACCTCAGAGATGGCGCTCATCCCTTGAT 479
Db      121 CTGGCAGACACACCTTCATTTTTCCTCAACCTCAGAGATGGCGCTCATCCCTTGAT 180
Qy      480 GCGCCACCATTCAGATGTCCTCCAGAGACCAAGCAACCTTTCCCGGAGCTCTCGACAC 539
Db      181 GCGCTACCATTCAGATGTCCTCCAGAGACCAACCTTTTCGAGAGCTCTCGACAC 240
Qy      540 GAAGTGTCAATTCATTCACCCCACTAGCCCACTCTGTACCGAGGAGTGTCTCC 599
Db      241 GAAGTGTCAATTCATTCACCCCACTAGCCCACTCTGTACCGAGGAGTGTCTCC 300
Qy      600 GAAGCAGAGAGAGTGTCTGCGGAGGAGACATGAGAGAAAGCTTGTGCGCGCGGAGAG 659
Db      301 GAAGCAGAGAGTGTGAGTGTGCGGAGGAGACATGAGAGAAAGCTTGTGCGCGGAGAG 360
Qy      660 CGCAACAGAGAGAGAGAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 719
Db      361 CGCAACAGAGAGAGAGAGTGTGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Qy      720 GCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 779
Db      421 GCCAATGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Qy      780 GGTGTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 839
Db      481 GGTGTCTGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540
Qy      840 GCCCAACATCATTTGGGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 899
Db      541 GCCCAACATCATTTGGGAGTGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 600
Qy      900 GAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 959
Db      601 GAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 660
Qy      960 GACTGGGAGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 1019
Db      661 GACTGGGAGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTAT 720
Qy      1020 TTGAGAGATTCCTGAGCTGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 1079
Db      721 TTGAGAGATTCCTGAGCTGAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAGTGCAGAG 780
Qy      1080 CTGGTGTCTGAGATCTCTTGGA 1103
Db      781 CTGGTGTCTGAGATCTCTTGGA 804

```

RESULT 9
AAC61089
ID AAC61089 standard: DNA; 5340 BP.

AAC61089;
AAC61089;
05-FEB-2001 (first entry)
Human neurogenin 3 (Ngn3) genomic DNA sequence.
Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
islet cell precursor identification; diabetes mellitus; human; ds.
Homo sapiens.
Key Location/Qualifiers
FH 3022..3666
FT CDS
FT /*tag= a

```

FT      /product= "Ngn3"
FT      /note= "Neurogenin 3"
FT      MO200059936-A1.
XX      12-OCT-2000.
XX      28-MAR-2000; 2000MO-US08436.
XX      06-APR-1999; 99US-0128180.
XX      (RBCG ) UNIV CALIFORNIA.
XX      German MS, Lin J;
XX      WPI; 2000-664989/64.
XX      P-PSDB; AAY85617.
XX      Novel human neurogenin 3 polypeptides and polynucleotides encoding
XX      them, useful for diagnosis, prevention and treatment of diabetes
XX      mellitus and to identify individuals at risk of diabetes -
XX      Claim 6; Page 46-48; 54pp; English.
XX      The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
XX      protein AAY85617. The Ngn3 gene is located at chromosome position
XX      10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
XX      protein sequences, and includes an antibody recognising the Ngn3 protein.
XX      Also included in the invention is a method for identifying an islet cell
XX      precursor, the method involves analysing a cell for the expression of the
XX      Ngn3 gene product, where detection of the product is indicative of an
XX      islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
XX      reagent for detecting (in a subject) a predisposition to a defect in
XX      pancreatic islet cell function or formation associated with a defect in
XX      Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
XX      precursor cells expressing Ngn3, and to alter cellular differentiation in
XX      culture in vivo to produce new beta-cells to treat patients with diabetes
XX      mellitus.
SQ      Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;

```

Query Match 35.4%; Score 517.4; DB 21; Length 5340;
Best Local Similarity 67.4%; Pred. No. 1.8e-130;
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;

```

Qy      20 GTCCCTGGGCCCCCGTTGCTATTTGGCCCGTGGAGAGAGAGAGAGAGAGAGAGAGAGAG 79
Db      2535 GAGCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2594
Qy      80 CCTGATCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 139
Db      2595 CTTGAGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2654
Qy      140 CCTCTGGGTCTACCACTGC-----ACAGAGCCGAGAGAGAGAGAGAGAGAGAG 181
Db      2655 TCCCTGGGTCTACCCGCGGCTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2714
Qy      182 CGAGCTTTTGTGCTGCTCCAGAGCAATTACT-CCAGAGAGAGAGAGAGAGAGAGAGAGAG 240
Db      2715 CTCTCTTTTCTTCTCTTTTGGGGCTGGGAGCACTCCAGAGAGAGAGAGAGAGAGAGAG 2774
Qy      241 CAAAGCTCGAAGGAGAGAG-----AGGGGTTCAGATTCACAGAGAGAGAGAGAGAGAG 285
Db      2775 CTGAAGCTTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2834
Qy      286 TGACTGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 345
Db      2835 TATCTTTTGGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2894
Qy      346 GGAAC-CTCCAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
Db      2895 AGAGGAGATCTGTGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2954

```

```

OY 405 GACCTTAAGTCAGAGACTG-----TCACACCCCTTCATTTTTCGCA 451
DB 2955 AGCCTCGAATGCGGAGCTGGCCCGGTGAACGACTCAACTTACCTTCCCTGACCCCG 3014
OY 452 CCTCAGATGCGGCTCATCTCTTGATGCGCCACCATCCAGTGTCTCCAAAGAACCA 511
DB 3015 CCGTAGATGACGCTCAACCTCCGGTGCCGCACTGTCCAAAGTACCCCGGAGACGGA 3074
OY 512 GCACCCCTTCCCGGAGCTCGGACCAAGATGCTCACTTCCATTCACCCCACTAG 571
DB 3075 GCGGTCTTCCCGAAGCTCGGAAAGAGATGACCTGCCCCAGTCCGCCGCCGCA 3134
OY 572 CCCCACTCTGATCCGAGGAGTGTCTCCAGACAGAGAGAGTGACTGCGGAGGACATC 631
DB 3135 CCCCACTCGACACAGGGGGGAACTGGCGCAGAGCGGAAAGAGAGAGGCTGCGAGGGGCC 3194
OY 632 GAGGAGCTCGGTGCGCGCGCGGAGAGGCGCAACAGGCCCAAGAGAGTGGCACTGAG 691
DB 3195 GAGGAGCTCCGGGCGACGGCGCGGGGAGCGCAGCCGCTTAAGAGCGAGTTGGCACTGAG 3254
OY 692 CAAGAGGAGGAGAGCCGGCGGCAAGAGGCGCAAGCGGGAGCGCAACCGCATGACAA 751
DB 3255 CAAGAGGAGGAGAGTGGCGGAAAGGCGCAAGCGCGAGCGCAATCGAATGACAA 3314
OY 752 CCTTAACCTCCGCGCTGATGCGCTGCGGAGTCTCTGCCCCACCTTCCGAGATGACGCCAA 811
DB 3315 CCTTAACCTCGACCTGAGACGCTTGGCGGTCTCTGCCCCACCTTCCGAGAGCGCGAA 3374
OY 812 ACTTAACAAGATCGAGACCTTGGCTTGGCCCAACATTAATTGGGCACTGACTGAGC 871
DB 3375 GCTCAACAAGATCGAGACGCTTGGCTTGGCCCAACATTAATTGGGCGCTGACTGACAA 3434
OY 872 GCTGGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 931
DB 3435 GCTGGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 931
OY 932 GGGAGAGCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 991
DB 3435 GCTGGGCAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 991
OY 992 AGCTGTAGCTGAGACCCCAAGCCTCATTTGAGAGAGTTCCCTGCGCTGACAGTGCAC 1051
DB 3555 GCGTGGAGCCTGAGATCCCGCGCGCTGCTGAGAGGAGCAGCCGGGCTGCTGGGAGC 3614
OY 1052 CTCCTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1111
DB 3615 CTCCTCCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3674
OY 1112 AACAGGCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1171
DB 3675 GCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3734
OY 1172 AAGGTAGTGAAGGCACTGAGCATC 1196
DB 3735 AAGGTAGTGAAGGCACTGAGCATC 1196

```

RESULT 10

```

ID AAD46871 standard; DNA; 5340 BP.
XX AAD46871;
AC AAD46871;
DT 27-JAN-2003 (first entry)
XX Human neurogenin 3 (Ngn3) gene.
DE Human neurogenin 3 (Ngn3) gene.
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KM type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KM islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
XX gene; ds.
OS Homo sapiens.

```

```

XX Key Location/Qualifiers
FH CDS 3022..3666
FT /+tag= a
FT /product= "Human Ngn3 protein"
PN WO200274045-A2.
XX 26-SEP-2002.
PF 20-MAR-2002; 2002WO-US11166.
XX 20-MAR-2001; 2001US-0817360.
PR (REGC ) UNIV CALIFORNIA.
PA German MS, Lin J;
PI WPI; 2002-759853/82.
XX P-PSDB; AAE29277.
DR Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor
XX Example 2; Page 87-88; 108pp; English.
CC The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 3
CC (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;
Query Match 35.4%; Score 517.4; DB 24; Length 5340;
Best Local Similarity 67.4%; Pred. No. 1.8e-130;
Matches 826; Conservative 0; Mismatches 351; Indels 48; Gaps 5;
OY 20 GTCCTGGGCCCCCGTGTGATTTGGCCGCGGACAGAGCCCGGAGGAGCGCT 79
DB 2535 GCGCCAGAGGCCCCCGGCTGATTTGGCCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 2594
OY 80 CTTGTGCGGAGAGAGATTAAGCTGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 139
DB 2595 CTTGCGCGGAGGAGGAGATTAAGCTGCCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2654
OY 140 CCTTGGGTCTACCACTGC-----ACAGAGCCGAGGAGAGCCCTC 181
DB 2655 TCCCTGCGGTCTCACCGCGCGGCTCGAGAGAGCGTGAAGAGGAGGAGGAGGAGGAG 2714
OY 182 CGAGCTCTTTGCTGCTCCAGAGCAATTAAT-CCAGGAGGAGGAGGAGGAGGAGGAGGAG 240
DB 2715 CTCCTCTTCTTCTCTTTGGGCTGCGGCACTCCAGGCGGAGGAGGAGGAGGAGGAG 2774
OY 241 CAATCTTGAAGCAGAG-AGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 285
DB 2775 CTGAACTTGGAGCAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2834
OY 286 TGACTCTGACACCGGAGCTCTGTTCTTTTGAAGCCCGAGATTAAGTAACTTTA 345
DB 2835 TATCTTTTGGCGCGGTAGAAAGATTAATTTTGAAGGAGGAGGAGGAGGAGGAGGAG 2894
OY 346 GGAAC-CTCCAAAGGTAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 404
DB 2895 AAGAGATCTCTGACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2954
OY 405 GACCTTAAGTCAGAGACTG-----TCACACCCCTTCATTTTTCGCA 451

```

```

Dh 2955 AGCCCTGGAAATCGCGGATCGCGCCGTGACGGAATCTAACTTACCTTCTCTAACCCTCG 3014
Qy 452 CCTCAGATGAGCGGCTCATCCCTTGAATGCCCAACCATCCAAATGTCCTCAAGAGACCA 511
Dh 3015 CCGAGATGACGCTCAACCTCGGTGCGCCCACTGTCCAAATGACCCGTGAGACGGA 3074
Qy 512 GCAACCTTTTCCGAGCTTGGACCAAGAGTGTCTCAATTCATCCCACTAG 571
Dh 3075 GCGGTCTTCTCCAGAGCTCGGAAGACGAATGACCTGCCCACTGCGCCCGCCAG 3134
Qy 572 CCCCACTCTGTACGAGGGAAGTCTCCGAGCAAGAGCAATGATCGCCGAGGACATC 631
Dh 3135 CCCCACTGTGACACAGGGGGAAGTGTGACAGCGGGAAGAGAGAGCTGCGCAGGAGCC 3194
Qy 632 GAGAAAGCTCCGTCCGCGCGGAGGAGCGCAAGAGCCCAAGAGCGATTGGCACTAG 691
Dh 3195 GAGAAAGCTCCGCGGACGCGCGGAGGAGCGAGCCGCTTAAGAGCGATTGGCACTAG 3254
Qy 692 CAAGCAGCGAGAAAGCCGCGCAAGAGGCAAGACCGGAGGCGCAACCGCATGCAAA 751
Dh 3255 CAAGCAGCGAGAGAGTGTGCGGAAAGAGGCAAGACCGGAGGCAATGCAATGCAAA 3314
Qy 752 CCTTAATCCGCGCTGATGCGCTGCGGCTGCTGCTCCCACTTCCGAGTGAAGCCAA 811
Dh 3315 CCTCAACTCGGACCTGAGCGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3374
Qy 812 ACTTAAAGATTCAGAGACCTGCGCTGCGGCAACATTCATTTGGGACATGATCAGAC 871
Dh 3375 GCTTACCAAGATTCAGAGACCTGCGCTGCGGCAACATTCATTTGGGACATGATCAGAC 3434
Qy 872 GCTGCGATAGCGAGACCAAGCTTCTACGCGCCGAGCCCGCTGTGCTGTGGGAGCT 931
Dh 3435 GCTGCGATAGCGAGACCAAGCTTCTACGCGCCGAGCCCGCTGTGCTGTGGGAGCT 3494
Qy 932 GGGAAAGCCCGGAGGAGGCTTCAAGGCGGACTGGGAGCTTCTTCTTCCCAATTTCCCA 991
Dh 3495 GCTGCGAGCGGAGGAGGCTTCCCGGAGGAGCTGGGAGGAGCTTCTTCTTCCCAATTTCCCA 3554
Qy 992 AGCTGTAGGCTGAGAGCCCAAGAGCTTCTTGAAGAGAGTTCCTGCGCGGAGGAGGAG 1051
Dh 3555 GCTGCGAGCGGAGGAGGCTTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3614
Qy 1052 CTCCCAATCTGTGCTGCTCCCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1111
Dh 3615 CTCTTCGCGCTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3674
Qy 1112 AACAGGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1171
Dh 3675 GCTGTGCGCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3734
Qy 1172 AAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1196
Dh 3735 AGGGTGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3794

```

RESULT 11
ABQ49522/c
ID ABQ49522 standard; DNA; 592 BP.

XX ABQ49522;

DT 12-JUL-2002 (first entry)

DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KM drug; side effect; cancer; central nervous system; cardiovascular;
KM gastrointestinal; respiratory system; single nucleotide polymorphism;
KM SNP; cell differentiation; ds.

XX Homo sapiens.

OS XX

```

Pn W0200218632-A2.
Xx
Pd 07-MAR-2002.
Xx
Pf 01-SEP-2001; 2001WO-EP10074.
Xx
Pr 01-SEP-2000; 2000DE-1043826.
Xx
Pr 05-SEP-2000; 2000DE-1044543.
Xx
Pa (EPID-) EPIGENOMICS AG.
Xx
Pi Olek A, Piepenbrock C, Berlin K, Guectig D;
Xx
Dr WPI; 2002-371829/40.
Xx
Pt Determining the degree of cytosine methylation in genomic DNA, useful
Pt for diagnosis and prognosis, comprises selective hybridization of
Pt amplicons from chemically treated DNA.
Xx
Ps Claim 12; 56pp + Sequence Listing; 56pp; German.
Xx
Cc This invention describes a novel method for determining the degree of
Cc methylation of a particular cytosine in a motif 5'-CpG-3', present in a
Cc genomic sample of DNA. The sample is treated chemically to convert
Cc cytosine (C) but not methylated C, to uracil, then part of the genomic
Cc DNA that contains the target C is amplified to form a labeled amplicon.
Cc The amplicon is hybridized to two classes, each with at least one
Cc member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
Cc and the degree of hybridization to both classes is determined from the
Cc label on the amplicon. From the ratio of labels hybridized to the two
Cc classes of oligomers, the degree of methylation is calculated. The method
Cc is used: (i) for diagnosis and/or prognosis of side effects of
Cc therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
Cc of the central nervous, cardiovascular, gastrointestinal and respiratory
Cc systems etc., particularly by detecting mutations or single nucleotide
Cc polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
Cc types and for investigating cell differentiation. The method allows the
Cc methylation status of many C residues to be determined simultaneously.
Cc AB013410-AB054121 represent genomic DNA sequences used to illustrate the
Cc method for determining the degree of cytosine methylation described in
Cc the disclosure of the invention.
Xx
S0 Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;
Xx
Qy Best Match 15.7%; Score 229.6; DB 24; Length 592;
Qy Best Local Similarity 67.6%; Pred. No. 2e-52;
Qy Matches 322; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
Dh 425 TCACACCCCTTCCATTTTTCACCACTCAGAGTGGCGCTCATCCTTGATGCGCC 484
Dh 477 TCACACTTACCTTCCCTCTTAACCCCGGTAATAAGGCTCAACCTCGAATATGCC 418
Qy 485 CACCATCAAGTGTCCCAAGAGACCAAGCAACCTTTCGGAGCTTCGACGAGAGT 544
Dh 417 CACTATCCAAATTAACCCGTAAACGAAGACATCTTCCCAAAACCTCGAAAAACGAAT 358
Qy 545 GCTCAGTTCAATTTCCACCCCACTAGCCCACTGTGTACCGGAGAGTGTCCGAGC 604
Dh 357 AACCTTACCCCAAGTGTGCGCCCGCCCAACCCCACTGCAACGAAAAAATCTAGCAAAAC 298
Qy 605 AGAAGCAGTGAAGTCCGAGGAGACATGAGGAAAGCTCGTCCGCGCGGAGGAGCCAA 664
Dh 297 GAAAAAATAAATACTACGAAAAAACCCGAAAAAATCTCGAACAGAGCGGAAAAACGCA 238
Qy 665 CAGGCCAAGAGCGAGTTGGCACTGACCAAGCAGCAAGGAGCCGCGCAAGAAAGCCAA 724
Dh 237 CCGACTTAAAAACCAATTAACATTAACAACAACGAGCAAGTGAAGAAAAAACCAG 178
Qy 725 CGACCGGAGAGCGCAAGCGCATGCAACATTAATCCGCGGTGAGTGGCGCGGTGT 784
Dh 177 CGACCGGAGAGCGCATGCAACATTAACATTAACATTAACATTAACATTAACATTAAT 118
Qy 785 CTTGCCACCTTCCCGAGTACGCGCAAACTTACAAAGATGAGAGCCTTGCGCTTGCCCA 844

```

Db 117 CCTACCCACCTTCCCAAGAGGAACTACCAAAATCGAAAGCTACGCTTGCCCA 58
 QY 845 CAATCACTTTGGGACCTGACTGACAGCGCTGGCTATAGGACCACTCTCTAG 900
 Db 57 CAATCACTTTGAAAGCTTAAGTCAAAAGCTACGATTAAGCAACCACTTATAG 2

RESULT 12

ABQ49523
 ID ABQ49523 standard; DNA, 592 BP.

ABQ49523;
 AC

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 drug; side effect; cancer; central nervous system; cardiovascular;
 gastrointestinal; respiratory system; single nucleotide polymorphism;
 SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 for diagnosis and prognosis, comprises selective hybridization of
 amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

CC This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
 CC genomic sample of DNA. The sample is treated chemically to convert
 CC cytosine (C) but not methylated C, to uracil, then part of the genomic
 CC DNA that contains the target C is amplified to form a labeled amplicon.
 CC The amplicon is hybridized to two classes, each with at least one
 CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
 CC and the degree of hybridization to both classes is determined from the
 CC label on the amplicon. From the ratio of labels hybridized to the two
 CC classes of oligomers, the degree of methylation is calculated. The method
 CC is used: (i) for diagnosis and/or prognosis of side effects of
 CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
 CC of the central nervous, cardiovascular, gastrointestinal and respiratory
 CC systems etc., particularly by detecting mutations or single nucleotide
 CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
 CC types and for investigating cell differentiation. The method allows the
 CC methylation status of many C residues to be determined simultaneously.
 CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
 CC method for determining the degree of cytosine methylation described in
 CC the disclosure of the invention.

XX Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Query Match 15.7%; Score 229.6; DB 24; Length 592;
 Best Local Similarity 67.6%; Pred. No. 2e-52;
 Matches 322; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 425 TCACACCCCTTCATTTTTCACACCTCAGAGTGGCGCTTCCTTGATGCC 484
 Db 116 TCAAACTTACCTTCCTCTTAACCCCGCGTAATATACGCTCAACCTCGAATAGGCC 175
 QY 485 CACCATCGAAGTGTCCCAAGAGACCAACCTTTCCCGAGACCTCGACACGAGT 544
 Db 176 CACTATCCAAATTAACCGTAAACGAAACGATCTTCCCAAAACCTGAAAGGAAAT 235
 QY 545 GCTCAGTTCCAAATTCACACCCACCTAGCCCACTCTGTACCGAGGACCTGCTCGAAGC 604
 Db 236 AACCTACCCACGTCGCGCCCGCCCAACCCCACTCGACACGAAAAAATAGCAAAAC 295
 QY 605 AGAAGCAGGTGACTGCGGAGGACATGAGAGCTCTGTCGCGCGGAGGCGCA 664
 Db 296 GAAAAAATAAACTAACGAAAAAATCCCGAAAAAATCGAAACAGACGCGAAAAACGAA 355
 QY 665 CAGGCCAAGAGCGAGTTGGACTGAGCAGACGACGAGAACCGCGGCGAAGAGCCAA 724
 Db 356 CCGACTTAAAAACGAATTTAACACTTAAACAAACGAGAAATGACGAAAAAATCCAA 415
 QY 725 CGACCGGAGGAGCAACGCGATGACAACTTAACCTCGCGCTGATGCGCTGCGGTGT 784
 Db 416 CGACCGGAGGAGCAATGAAATACACACTTAACCTGACACTTAAACGCTTACCGGATAT 475
 QY 785 CCGTCCCACTTCCCGGATGACGCGCAAACTTAAGATGAGACCTGCGCTTGCCTCA 844
 Db 476 CCTACCCACCTTCCCAAGAGGAACTACCAAAATCGAAAGCTTACGCTTGCCTCA 535
 QY 845 CAATCACTTTGGGACCTGACTGACAGCGCTGGCATAGCGGACCAAGCTTCTAG 900
 Db 536 CAATCACTTTGAAAGCTTAAGTCAAAAGCTACGATTAAGCAACCACTTATAG 591

RESULT 13

ABQ49524
 ID ABQ49524 standard; DNA, 592 BP.

ABQ49524;
 AC

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 36115.

Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
 drug; side effect; cancer; central nervous system; cardiovascular;
 gastrointestinal; respiratory system; single nucleotide polymorphism;
 SNP; cell differentiation; ds.

OS Homo sapiens.

PN WO200218632-A2.

PD 07-MAR-2002.

PF 01-SEP-2001; 2001WO-EP10074.

PR 01-SEP-2000; 2000DE-1043826.

PR 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX WPI; 2002-371829/40.

PT Determining the degree of cytosine methylation in genomic DNA, useful
 for diagnosis and prognosis, comprises selective hybridization of
 amplicons from chemically treated DNA -

PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
 CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a

genomic sample of DNA. The sample is treated chemically to convert cytosine (C) but not methylated C, to uracil, then part of the genomic DNA that contains the target C is amplified to form a labeled amplicon. The amplicon is hybridized to two classes, each with at least one member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers and the degree of hybridization to both classes is determined from the label on the amplicon. From the ratio of labels hybridized to the two classes of oligomers, the degree of methylation is calculated. The method is used: (i) for diagnosis and/or prognosis of side effects of therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders of the central nervous, cardiovascular, gastrointestinal and respiratory systems etc., particularly by detecting mutations or single nucleotide polymorphisms (SNP's); and (ii) for differentiation of cell or tissue types and for investigating cell differentiation. The method allows the methylation status of many C residues to be determined simultaneously. AB013410-AB054121 represent genomic DNA sequences used to illustrate the method for determining the degree of cytosine methylation described in the disclosure of the invention.

Sequence 592 BP; 123 A; 59 C; 187 G; 223 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;

Best Local Similarity 64.5%; Pred. No. 3.1e-42;

Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
OY 456 AGATGGGCGCTCATCCCTTGATGCGCCACATCCAGTGTCCCAAGACCAGCA 515
    |||||
DB 147 AGAGATGAGTTTAAATTTTCGGGTGCTTTATTTAGTATTCGAGAGCGAGGG 206
OY 516 CCGTTTCCCGAGCTCGGACCAAGTGTCTCAATTCACCCACCTTACCCC 575
    |||||
DB 207 TTTTATTTAGAGTTTCCGAAAGACGAGTGTATTTTACCTTCTGTTTCTTT 266
OY 576 ACTCCGATACGAGGAGCTGCTCGAAGACAGAGTGTGCGAGGGACATCGAG 635
    |||||
DB 267 ATTCTATACGAGGAAATTCGTAGAGCGGAGGAGGAGGTTGTCAGGGGTTTCGAG 326
OY 636 AAGCTCCGTGCGCGCGGAGGAGGCGCAACAGGCGCAAGAGGAGTGTGCACTGAGCAAG 695
    |||||
DB 327 AAGTTTCGGGTACGCGCGCGGAGAGTGTGCTTTAAGAGGAGTGTGATGATG 386
OY 696 CAGCGACGAGCGCGGCAAGAGGCGCAACGAGCGGAGCGGACCGATGACAACTT 755
    |||||
DB 387 TAGCGACGAGTGTGCGGAAAGAGTAAAGATCGAGCGATGATGATATTTT 446
OY 756 AACTCCGCGCTGAGTGGCTGTGCGGTCTCTGCCACCTTCCCGATGAGCGCAACTT 815
    |||||
DB 447 AATTCGATTTGAGAGGTTTTCGCGGTGTGTTTATTTTATTTTGAAGACGCGAAGTTT 506
OY 816 ACAAGATCGAGACCTGCGGCTTCCGCCAACAATCTATTTGGGACACTGACAGCGCTG 875
    |||||
DB 507 ATTAAAGATCGAGAGCTTGGCTTCTTTATTAATTAATTTGGGCGTTATTTAAACGTTG 566
OY 876 CGCATAGCGGACCAAGCTTTACG 900
    |||||
DB 567 CGTATAGCGGATTAATGTTTACG 591
    |||||
```

RESULT 14

AB049525/C

AB049525 standard; DNA; 592 BP.

AB049525;

12-JUL-2002 (first entry)

Oligonucleotide for detecting cytosine methylation SEQ ID NO 36116.
Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
drug; side effect; cancer; central nervous system; cardiovascular;
gastrointestinal; respiratory system; single nucleotide polymorphism;
SNP; cell differentiation; ds.

OS Homo sapiens.
XX
XX WO200218632-A2.
XX
XX 07-MAR-2002.
XX
XX 01-SEP-2001; 2001WO-EPI0074.
XX
XX 01-SEP-2000; 2000DE-1043826.
PR 05-SEP-2000; 2000DE-1044543.
XX
XX (EPIG-) EPIGENOMICS AG.
XX
XX Olek A, Piepenbrock C, Berlin K, Guetig D;
PI WPI; 2002-371829/40.
DR
XX

Determining the degree of cytosine methylation in genomic DNA, useful
for diagnosis and prognosis, comprises selective hybridization of
amplicons from chemically treated DNA

Claim 12; 56pp + Sequence Listing; 56pp; German.

This invention describes a novel method for determining the degree of
methylation of a particular cytosine in a motif 5'-CpG-3', present in a
genomic sample of DNA. The sample is treated chemically to convert
cytosine (C) but not methylated C, to uracil, then part of the genomic
DNA that contains the target C is amplified to form a labeled amplicon.
The amplicon is hybridized to two classes, each with at least one
member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
and the degree of hybridization to both classes is determined from the
label on the amplicon. From the ratio of labels hybridized to the two
classes of oligomers, the degree of methylation is calculated. The method
is used: (i) for diagnosis and/or prognosis of side effects of
therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
of the central nervous, cardiovascular, gastrointestinal and respiratory
systems etc., particularly by detecting mutations or single nucleotide
polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
types and for investigating cell differentiation. The method allows the
methylation status of many C residues to be determined simultaneously.
AB013410-AB054121 represent genomic DNA sequences used to illustrate the
method for determining the degree of cytosine methylation described in
the disclosure of the invention.

Sequence 592 BP; 223 A; 187 C; 59 G; 123 T; 0 other;

Query Match 13.2%; Score 192.2; DB 24; Length 592;

Best Local Similarity 64.5%; Pred. No. 3.1e-42;

Matches 287; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

```
OY 456 AGATGGGCGCTCATCCCTTGATGCGCCACATCCAGTGTCCCAAGACCAGCA 515
    |||||
DB 446 AGAGATGAGTTTAAATTTTCGGGTGCTTTATTTAGTATTCGAGAGCGAGGG 387
OY 516 CCGTTTCCCGAGCTCGGACCAAGTGTCTCAATTCACCCACCTTACCCC 575
    |||||
DB 386 TTTTATTTTAAAGTTTGGGAAGAGAGTATTTGTTTACGTTCTGTTTAACTT 327
OY 576 ACTCTGATCCGAGGAGCTGCTCGAAGACAGAGCGTGAATCCGAGGACATCGAG 635
    |||||
DB 326 ATTGATATACGAGGAAATTCGTAGAGCGGAGGAGGAGGTTTTCGAGGGGTTTCGAG 267
OY 636 AAGCTCCGTGCGCGCGGAGGAGGCGCAACAGGCGCAAGAGGAGTGTGCACTGAGCAAG 695
    |||||
DB 266 AAGTTTCGGGTACGCGCGCGGAGAGTGTGCTTTAAGAGCGAGTGTGATTAAGTAA 207
OY 696 CAGCGACGAGCGCGGCAAGAGGCGCAACGAGCGGAGCGCAACCGATGACCAACTT 755
    |||||
DB 206 TAGGACGAGAGTGTGCGGCAAGAGGATTAAGATGCGAGGTATTAATTAATTTT 147
OY 756 AACTCCGCGCTGAGTGGCTGTGCGGTCTCTGCCACCTTCCCGATGAGCGCAACTT 815
    |||||
DB 146 AATTCGATTTGAGAGCTTTCGCGGTGTGTTTATTTTATTTTGAAGACGAGCAAGTTT 87
    |||||
```

QY 816 ACAAGATCGAGACCCCTGCGCTTCCGCCACACTACATTTGGGCACTGACTCAGACGCTG 875
 DB 86 ATTAAGATCGAGACCTTGCGCTTCCGTTTAATTAATTTGGGCGGTGATTTAAACGTTG 27
 QY 876 CGCATAGCGGACACACAGCTTCTACG 900
 DB 26 CGTATAGCGGATTAATGTTTGTACG 2

RESULT 15
 ID AAD46889
 AAD46889 standard; DNA; 714 BP.
 AC AAD46889;
 DT 27-JAN-2003 (first entry)
 XX Human neurogenin 1 (Ngn1) gene #2.
 DE Human neurogenin 1 (Ngn1) gene #2.
 XX
 KM Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
 KM type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
 KM islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
 XX
 OS Homo sapiens.
 XX
 FT Key Location/Qualifiers
 FT CDS 1..714
 FT /tag= a
 FT /product= "Human Ngn1 protein"
 XX
 W0200274045-A2.
 XX 26-SEP-2002.
 PD
 XX 20-MAR-2002; 2002WO-US11166.
 PF
 XX 20-MAR-2001; 2001US-0817360.
 PR
 XX (REGC) UNIV CALIFORNIA.
 PA
 PI German MS, Lin J;
 XX
 DR WPI: 2002-759853/82.
 DR P-PSDB; AAE29280.
 DR
 XX
 XX
 PT Producing a mammalian islet cell for treating diabetes mellitus
 PT comprises introducing into a mammalian cell a nucleic acid molecule
 PT encoding neuroendocrine basic helix-loop-helix transcription factor -
 XX
 PS Disclosure; Page 94; 108pp; English.
 XX
 CC The invention relates to a method for producing a mammalian islet cell.
 CC The method comprising introducing into a mammalian cell a nucleic acid
 CC molecule encoding an islet transcription factor for expression of the
 CC islet transcription factor in the cell and for production of islet cell
 CC phenotype in the cell. The islet transcription factor is a neuroendocrine
 CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
 CC for treating type 2 diabetes mellitus and for replacing beta cells lost
 CC to autoimmune destruction in individuals with type 1 diabetes. The method
 CC is useful in cell therapy. The present sequence is human neurogenin 1
 CC (Ngn1) gene.
 CC
 SQ Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;

Query Match 10.1%; Score 147.6; DB 24; Length 714;
 Best Local Similarity 71.2%; Pred. No. 4,7e-30;
 Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGGAGAGGCGGCAACAGGCCCAAGAGGAGTTGGCACTGAGCAAGCAGCAGAAAG 706
 DB 216 GCGGCGGCGGCGGCGGAGCGGAGGTCGCTCCGAGCGCGCTGCTGCACTCGCTGCGCAGAG 275

QY 707 CCGGCGCAAGAGGCCAAGCAGACCGGAGCGCAACCGCATGCAAACTTAACTCCGCGCT 766
 DB 276 CCGGCGGCTCAAGGCCAAGCATCGCGAGCGCAACCGCATGCAAACTTGAACGGGCGCT 335
 QY 767 GGAATGCGGTGCGGAGTGTCTGCTGCCACCTTCCGGAATGAGCCCAAACTTAAGAAATGA 826
 DB 336 GGAAGCACTGCGCAGCGGTGCTGCCCTCGTTCCCGAGACACCAAGCTCACCAAAATGA 395
 QY 827 GACCCGTGCTTGGCCCAACTACATTTGGGCACTGACTGAGCGCTGCGCATAGCGGA 886
 DB 396 GACGTGCGCTTGGCTTACAACTACATCTGAGGCTTGGCGGAGACACTGCGCTGCGGA 455
 QY 887 CCACAGCTTCTACGCGCCCGAGGCCCGCTGTGCCC 920
 DB 456 TCAAGGCGTGCCTCGGAGGCGGTGCCCGGAGCGC 489

Search completed: January 26, 2004, 19:39:01
 Job time : 430 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 26, 2004, 18:48:21 ; Search time 98 Seconds
(without alignments)
6575.708 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcaggtacgagagagagagcag.....agagtacctaaccagtc 1460

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTOS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	652	44.7	804	US-08-932-411A-19	Sequence 19, Appl
2	147.6	10.1	1268	US-08-910-973-12	Sequence 12, Appl
3	147.6	10.1	1268	US-09-499-227-12	Sequence 12, Appl
4	145.2	9.9	1527	US-08-722-570-12	Sequence 12, Appl
5	145.2	9.9	1527	US-08-932-411A-12	Sequence 12, Appl
6	140.4	9.6	738	US-08-722-570-13	Sequence 13, Appl
7	140.4	9.6	738	US-08-932-411A-13	Sequence 13, Appl
8	140.4	9.6	1333	US-08-910-973-21	Sequence 21, Appl
9	140.4	9.6	1333	US-09-499-227-21	Sequence 21, Appl
10	139.8	9.6	1385	US-08-932-411A-17	Sequence 17, Appl
11	121.8	8.3	310	US-08-552-142A-12	Sequence 12, Appl
12	115.6	7.9	1312	US-08-722-570-14	Sequence 14, Appl
13	115.6	7.9	1312	US-08-932-411A-15	Sequence 15, Appl
14	109.8	7.5	1277	US-08-722-570-15	Sequence 15, Appl
15	109.8	7.5	1277	US-08-932-411A-16	Sequence 16, Appl
16	96	6.6	1352	US-08-552-142A-10	Sequence 10, Appl
17	96	6.6	1352	US-08-910-973-10	Sequence 10, Appl
18	96	6.6	1535	US-09-499-227-10	Sequence 10, Appl
19	96	6.6	1550	US-09-234-332-3	Sequence 3, Appl
20	93.2	6.4	1462	US-08-552-142A-16	Sequence 16, Appl
21	93.2	6.4	1951	US-08-910-973-16	Sequence 16, Appl
22	93.2	6.4	1951	US-09-499-227-16	Sequence 16, Appl
23	92.4	6.3	2089	US-08-552-142A-1	Sequence 1, Appl
24	92.4	6.3	2089	US-08-910-973-1	Sequence 1, Appl
25	92.4	6.3	2089	US-09-499-227-1	Sequence 1, Appl
26	92.4	6.3	2089	PCT-US95-05741-1	Sequence 1, Appl
27	89	6.1	1560	US-08-552-142A-14	Sequence 14, Appl

28	89	6.1	1560	1	US-08-910-973-14	Sequence 14, Appl
29	89	6.1	1560	4	US-09-499-227-14	Sequence 14, Appl
30	89	6.1	2502	3	US-09-234-332-1	Sequence 3, Appl
31	85.8	5.9	1676	3	US-08-234-332-2	Sequence 2, Appl
32	83.4	5.7	524	1	US-08-552-142A-8	Sequence 8, Appl
33	83.4	5.7	524	1	US-08-910-973-8	Sequence 8, Appl
34	83.4	5.7	524	4	US-09-499-227-8	Sequence 8, Appl
35	83.4	5.7	524	5	PCT-US95-05741-8	Sequence 8, Appl
36	80	5.5	485	5	PCT-US95-05741-10	Sequence 10, Appl
37	73.6	5.0	1275	1	US-08-552-142A-3	Sequence 3, Appl
38	73.6	5.0	1275	1	US-08-910-973-3	Sequence 3, Appl
39	73.6	5.0	1275	5	US-09-499-227-3	Sequence 3, Appl
40	73.6	5.0	1275	5	PCT-US95-05741-3	Sequence 3, Appl
41	51.6	3.5	1635	3	US-09-234-332-4	Sequence 4, Appl
42	51.6	3.5	1635	4	US-09-702-705-1798	Sequence 1798, Ap
43	51.6	3.5	1635	4	US-09-736-457-1798	Sequence 1798, Ap
44	41	2.8	1785	5	PCT-US94-12912-1	Sequence 1, Appl
45	41	2.8	3636	1	US-07-753-520B-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-932-411A-19
; Sequence 19, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 804 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 160..801
; US-08-932-411A-19

Query Match	44.7%;	Score 652;	DB 4;	Length 804;
Best Local Similarity	90.9%;	Pred. No. 4.6e-163;		
Matches 731; Conservative	0;	Mismatches 60;	Indels 13;	Gaps 3

QY	312	TTCTTTTAAAGCCGGAGTAACATGAGTAACATTTAGAACTCCAAAGGGTATGAAGGGG	371
Db	2	TTCTTTTAAAGTCGGAG-A-ACATAGTAACATTCGMAAATCCMAAGGTGATGAGGGG	60
QY	372	AGTGGGT-----GGGCGTACTAGTCCGCGTGAGTACCTCTAAGTCAGAGA	421
Db	61	CGCGCGGGGTGTGTGTGGGGATATCTGTGTCCCGTCGTACATGACTCTAAGTCAGAGG	120
QY	422	CTG--TCAACCCCCCTTCATTTTTCCTCCAACTCAGATGGCGCTCATCCCTTGAT	479
Db	121	CTGGCACACACACACCTTCATTTTTCCTCCAAACGAGATGAGCGCTCATCCCTTGAT	180
QY	480	GCGCCACCATATCAGTGTCCCAAGAGACCAGCAACCTTTCGCGAGCTTCGACAC	539
Db	181	GCCTCAACCATTCAGTGTCCCAAGAGACAAACAACCTTTCGCGAGCTTCGACAC	240
QY	540	GAAGTCTCAGTTCCATTTCCACCCCACTAGCCCACTCTGTATCCAGGGACTGTCTCC	599
Db	241	GAAGTCTCAGTTCCATTTCCACCCCACTAGCCCACTCTCATCTAGGGACTGTCTCC	300
QY	600	GAAAGAGAAAGACAGTGACTGCCGAAGGACATGAGGAACTCCGTGCGGGCGGAGGG	659
Db	301	GAAAGAGAAAGTGGTGACTGCCGAAGGACCTGAGGAACTCCGTCCGACGCGGAGGG	360
QY	660	CGCAACAGAGCCCAAGAGGAGTTTGGCACTGAGCAAGCAGACGAAAGCCGGCGCAAGAAG	719
Db	361	CGCAACAGAGCCCAAGAGGAGTTTGGCACTGAGCAACAGCAGAAAGCCGGCGCAAGAAG	420
QY	720	GCCAAAGACCGGGAGCGCAACCGCATGACAAACCTTAACTCCGCGTGGATGCGCTGGCG	779
Db	421	GCCAAATGATCGGGAGCGCAATGCAATGCAACCTTCAACTCGCGCGTGGATGCGCTGGCG	480
QY	780	GATGTCTCGGCCACCTTCGCCGATGACGCCAAACTTAAGATGACAGACCTTGCGCTTC	839
Db	481	GATGTCTCGGCCACCTTCGCCGATGACGCCAAACTTAAGATGACAGACCTTGCGCTTC	540
QY	840	GCCCAACAATCATTTTGGGCACTGACTCAGACGCTGCGCATGACGGAACCAAGCTTCTAC	899
Db	541	GCCCAACAATCATTTTGGGCACTGACTCAGACGCTGCGCATGACGGAACCAAGCTTCTAC	600
QY	900	GAGCCCGAGCCCCCTGTGTGCTGTGTGGGAGCTGGGAAGCCCGGAGAGGGGGCTTCACGGC	959
Db	601	GAGCCCGAGCCCCCTGTGTGCTGTGTGGGAGCTGGGAAGCCCGGAGAGGGGGCTTCACGGG	660
QY	960	GACTGGAGCTCATATCATCCCCCAGTTTCCAAAGCTGTAGCTGAGCCCAAGGCTCTGA	1011
Db	661	GACTGGAGCTCATATCATCCCCCAGTTTCCAAAGGCTGTAGCTGAGCCCAAGGCTCTGA	720
QY	1020	TTGAGAGAGTTCCTTGAGCTCTGAGGTGCCAGCTCCCATCTGTCTGTCCCGGGACCC	107
Db	721	TTGAGAGAGTTCCTTGAGCTCTGAGGTGCCAGCTCCCATCTGTCTGTCCCGGGAGCA	780
QY	1080	CTGTGTCTCTCAGACTTCTTGGA 1103	
Db	781	CTGTGTCTCTCAGACTTCTTGGA 804	

RESULT 2
US-08-910-973-12

TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectodermal
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC

STREET 1420 Fifth Avenue, Suite 2800
City: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,973
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FPCR-1-10958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1268 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: 20A1 (neurod3)
FEATURE:
NAME/KEY: CDS
LOCATION: 55..768

	Query Match	10.1%;	Score 147.6;	DB 1;	Length 1268;
	Best Local Similarity	71.2%;	Pred. No. 1.1e-29;		
	Matches 195;	Conservative 0;	Mismatches 79;	Indels 0;	Gaps 0;
QY	647	GCAGCGCGAGAGGCGCAACAGGCGCCAGAGCGATTGGCACTGAGCAAGCAGCGACGAAG	706		
Dh	270	GCAGCGCGCGCGCGCGAGCGCGGATGCCCTCCGAGAGCGCGTGTCTGCACTCGTCGCGAGGAG	329		
QY	707	CCGGCGCGAAGAGGCCAAGCGACGGGAGCGCGAACCGCATGACAAACCTTAATCCGCGCT	766		
Dh	330	CCGGCGCGGTCAAGGCGCAACGATCGCGAGCGCAACCGCATGACAACTTGAACGGCGCCT	389		
QY	767	GGATGCGCTGCGCGCGTGTCTCGACCCACCTTCCCGATGACCGCAACCTTACAAAGATGCA	826		
Dh	390	GGAGCGACTGCGCGACGCTGTCTGCTCCCTCGTTCCCGAGACGACCAAGCTCACCAAAATGCA	449		
QY	827	GACCTCGCGCTTGGCCCAACATTCATTGGGCACTGACTCAGACGCTGGCGATAGCGGA	886		
Dh	450	GACGCTGGCTTTCGCTTACAACTACATCTGGGCTCTGGCCGAGACACTGCGCCTGGCGCGGA	509		
QY	887	CCACAGCTTTCAGCGCGCCCGGAGCCCGCTGTGCC	920		
Dh	510	TCAAGGCGTCCCGAGAGCGGTGTCCCGGAGCGC	543		

Sequence 12, Application US/09499227
Patent No. 644463
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09499,227
FILING DATE: 05-August-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-May-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-May-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910,973
FILING DATE: 07-August-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sheinnes, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FPCR-1-12742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1268 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: 20A1 (neuroD3)
FEATURE:
NAME/KEY: CDS
LOCATION: 55..768
US-09-499-227-12

Query Match 10.1%; Score 147.6; DB 4; Length 1268;
Best Local Similarity 71.2%; Pred. No. 1.1e-29;
Matches 195; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGGGCGGAGCGGCAAGCGCCCAAGAGGAGTTGACATGACAGCAAGCGAGCAAG 706
DB 270 GCGGGGCGGAGCGGCGGAGCGGAGTTCCTCGAGGCGCTGCACTGCGCGAGAG 329
QY 707 CCGGCGAAGAGCGCAAGCGGAGCGCAACCGCATGCAACAATTAACTCCGCGCT 766
DB 330 CCGGCGGCTCAAGCGCAAGCATCGGAGCGCAACCGCATGCAACAATTAACTCCGCGCT 389
QY 767 GGATGCGTGGCGGATGCTGCGCACTTCCCGGATGACGCAAACTTACAAATGCA 826
DB 390 GGAGCGACTGCGCAAGCTGCTGCTCTGTTCCCGAGCAACCAAGCTCACCAATGCA 449

QY 827 GACCTGCGCTTGCCCAACAATTGAGCACTGACTCAGACGCTGGCATAGGGA 886
DB 450 GACGCTGCTTCCGCTTCAACAATCTGAGGCTGCTGCGAGCACTGCGCTGGCGGA 509
QY 887 CCACAGCTTCAAGCGCCGAGCGCCGCTGTGCCC 920
DB 510 TCAAGGCTGCGGAGCGGCTGCTGCGGAGCGC 543

RESULT 4
US-08-722-570-12
Sequence 12, Application US/08722570
Patent No. 655337
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Teet, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1527 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-12

Query Match 9.9%; Score 145.2; DB 4; Length 1527;
Best Local Similarity 69.2%; Pred. No. 4.8e-29;
Matches 198; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 635 GAAGCTCGTGGGCGGCGGAGCGGCAAGCGCCCAAGAGTGGGCACTGAGCA 694
DB 456 GCAGAGCGGCGGCGGAGCGGAGTTCGCGCGGCTGCGATCCAGAGCGCTGCACTC 515
QY 695 GCAAGCAAGAGCGGCGGCAAGAGCGCAAGCGGAGCGCAACCGCATGCAACAACCT 754
DB 516 GCTGCGGAGGAGCGGCGGCAAGAGCGCAAGCGGAGCGCAACCGTATGCAATACT 575
QY 755 TAACTCGCGCTGATGATGCTGCGCGGCTGCTGCGCAACTTCCCGGATGACGCAACT 814
DB 576 CAAAGCTGCGTGAAGCTTGCAGAGTGTCTCTGTTCCCGAGCAACCAAGCT 635
QY 815 TACAAGATGAGAGCCGTGCGCTTGCAGCAACTAATTGGGCACTGACTAGACGCT 874
DB 636 CACCAAGATTAGAGCGCTGCGCTTGCCTTACATCACTTGGGCGCTGAGACACT 695
QY 875 GCGCATAGCGGACCAAGCTTCTACGCGCCCGAGCGCCCTGTGCCC 920

Patent No. 6566496
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qulu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..732
US-08-932-411A-13

	Query Match	Similarity	9.6%	Score 140.4	DB 4	Length 738
	Best Local	Similarity	68.2%	Pred. No. 7.2e-28		
	Matches	195	Conservative	0	Mismatches 91	Indels 0
QY	635	GAAGCTCGGTGCGCGGCGCGGAGGGGCGCAACAGGCCCAAGAGAGGAGTTGGACTGTAGCA	694			
Db	207	GCAGGAACGGCGCGAGCGCGCGCGAGGTGCGCCTCGGGGCGAGTCCGAGGCTCTGCTGCACTC	266			
QY	695	GCAGGACGAAAGCCGCGCGCAAGAGAGGCGCAACGACCGGAGCGCAACCGGATGCACAACT	754			
Db	267	CCTGGGAGGAGATCGTTCGCTCAAAAGCCAAAGATCGGAGCGCGAACCGGATGCACAACT	326			
QY	755	TAACTCCGCGCTGGATGCGCTGCGGGGTCTCTGCCACCTTCCGGATGACGCCAACT	814			
Db	327	CAGCCCTCGCTGGACGCGCTTGCAGACGTCGTGCGCTCGTTCGCCCGAGAACCAAGCT	386			
QY	815	TACAAAGATCGAGACCCCTGCGCTTGCAGCCACAACTATTGGGCACTGACTCAGACGCT	874			
Db	387	CACCAAGATTGAGAGGCTGCGCTTGGCTTCAACAATCACTGGGCGCTTGGCTGAGCACT	446			
QY	875	GCGCATAGCGACCAACAGCTTTACGAGGCCCGAGGCCCTCTGTGCC	920			
Db	447	GCGCTGGAGATCAAGGCGCTCCCGGGGGCAGTGGCCCGGAGCGC	492			

```

RESULT 8
US-08-910-973-21
Sequence 21, Application US/08910973
Patent No. 5795723
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESS: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,973
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHC-1-10958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Mus musculus
IMMEDIATE SOURCE:
CLONE: neuroD3
FEATURE:
NAME/KEY: CDS
LOCATION: 101..835
US-08-910-973-21

```

	Only Match	9.6%	Score 140.4	DB 1	Length 133
	Best Local Similarity	68.2%	Pred. No. 8.6e-26		
	Matches 195	Conservative	0	Mismatches 91	Indels 0
				Gaps	0
Qy	635	GAAGCTCCGTGGCGCGCGAGGCGCGCAAGGCCCAAGAGCGAGTTGGCACTGAGCAA	694		
Db	307	GCAGGAACGGCGGAGAGGCGGCGAGGTCGGGTCGGGTGCGGTCCGAGGCTCGCTGCACCTC	366		
Qy	695	GCAGCGAAGACCCGCGCGAAGAGGCCAAGACCCGAGGCGCAACCGCATGCAACAACTT	754		
Db	367	CTGCGGAGAAATCGTCGGGTCAAGCAACGATCGAGGCGCAACCGCATGACAACTT	426		
Qy	755	TAACTCCGCGCGTGAATGGCTGCGGGGTCTCTGCCCACTTCCCGGATATACCCAAACTT	814		
Db	427	CAACGCTGCGCTGAAGCCCTTCGCAAGAGGTCTGCGCTTCCCGAGCAACCAAGCT	486		

Oy 815 TACAAAGATGAAACCCCTGGCTGGCCGACAACTACATTGGGACATGATCAGACGCT 874
Db 487 CACCAAGATTGAAACGCTGGCTGGCCCTCAACAATCAATCTGGGCGCTGGGCTGAGACACT 546
Oy 875 GCGCATGAGCGAACCAACAGCTTTATGAGGCCCGGAGCCCTCTGTGCCCC 920
Db 547 GCGCTGGAGAGATCAAGGGCTCCCGCGGGAGCAAGTGGCCGGAGAGCGC 592

RESULT 9

```

Sequence 21, Application US/093499227
Patent No. 6444463
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olsson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoderm
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Christensen O'Connor Johnson KindnessPLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle

```

```

1      ZIP: 98101-2347
2      COMPUTER READABLE FORM:
3      MEDIUM TYPE: Floppy disk
4      COMPUTER: IBM PC compatible
5      OPERATING SYSTEM: PC-DOS/MS-DOS
6      SOFTWARE: Patent Release #1.0, Version #1.25
7      CURRENT APPLICATION DATA:
8      APPLICATION NUMBER: US/09/499,227
9      FILING DATE: 05-August-1998
10     PRIOR APPLICATION DATA:
11     APPLICATION NUMBER: US 08/239,238
12     FILING DATE: 06-May-1994
13     PRIOR APPLICATION DATA:
14     APPLICATION NUMBER: WO PCT/US95/05741
15     FILING DATE: 08-May-1995
16     PRIOR APPLICATION DATA:
17     APPLICATION NUMBER: PCT/US96/17512
18     FILING DATE: 30-October-1996
19     PRIOR APPLICATION DATA:
20     APPLICATION NUMBER: US 08/910,973
21     FILING DATE: 07-August-1997
22     ATTORNEY/AGENT INFORMATION:
23     NAME: Sheinesa, Diana K.
24     REGISTRATION NUMBER: 35,356
25     REFERENCE/DOCKET NUMBER: FPCR-1-12742
26     TELECOMMUNICATION INFORMATION:
27     TELEPHONE: 206-662-8100; 206-224-0735 (direct)
28     TELEFAX: 206-225-0779
29     INFORMATION FOR SEQ ID NO: 21:
30     SEQUENCE CHARACTERISTICS:
31     LENGTH: 1333 base pairs
32     TYPE: nucleic acid
33     STRANDEDNESS: double
34     TOPOLOGY: linear
35     MOLECULE TYPE: DNA (genomic)
36     ORIGINAL SOURCE:
37     ORGANISM: Mus musculus
38     IMMEDIATE SOURCE:
39     CLONE: neurod3
40     FEATURE:
41     NAME/KEY: CDS
42     LOCATION: 101..835
43     US-09-499-227-21

```

Query	Match	Score	DB	Length
635	GAAAGCTCCGTTGGCGCGCGCGGAGGCGGCGCAACAGCCCAAGAGCGAATTGGCACTAGACCA	140.4	4	1333
	Best Local Similarity	68.2%	Pred. No. 8	6e-28
	Matches 195, Conservative	0	Mismatches 91	Indels 0, Gaps 0

Db	307	GCAGGAACGGCGGAGGGCGGAGGTCGTGGGTGCGGTCCGAGCTCTGCTGCACCTC	366
Oy	695	GCAGCGACGAACCCGGCGCAGAAAGGCGCAACGCGCAACGATGCACAACT	754
Db	367	CTGCGGAGAGATCTGTGCTGTAAGACCAAGATGCGAGCGCAACCGCATGCACAACT	426
Oy	755	TAACTCCGCTGGATGCGCTGCGCGGTGTCCTGCGCACTTCCCGATGACGCAAACT	814
Db	427	CAACGTGGCGCTGGACGCTTCGCGAGGATCTGTGCTCTCCCGACGACAAAGCT	486
Oy	815	TACAAAGATCGAGACCTCGCTCTGGCCACAACTATTTGGGCACTGACTCAGAGCT	874
Db	487	CACCAAGATTGAGACGTGCTGTGGCTTAACAATCTGGGCCCTGGCTGAGACACT	546
Oy	875	GGCGATAGCGGACCAACGTTTAAAGGCCCGGAGCCCTGTGGCC	920
Db	547	GGCGCTGGCAGATCAAGGGCTCCCGGGGGGACGTGCCCGGAGAGGC	592

RESULT 10

```

US-08-932-411A-17
Sequence 17, Application US/08932411A
Patent No. 6566496
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSER: Flahir Hohbach Test Albritton & Herdert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1385 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:

```

```

; LOCATION: 382..1170
US-08-932-411A-17

Query Match          9.6%; Score 139.8; DB 4; Length 1385
Best Local Similarity 78.1%; Pred. No. 1.3e-27;

```

Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

Qy 705 AGCCGCGCAGAAAGGCAAGAGCGAGGAGCAACCGCATGCAACCTTAACCTCCGCG 764
Db 715 ACCCGCAGGCTCAAGAGCCCAACCGGAGCAACCGCATGCAACCTTAACCGCGCG 774
Qy 765 CTGATGCGCTGCGCGGCTGCTGCTGCACTTCCCGATGACGCCAACTTAAGATC 824
Db 775 CTGACGCGCTGCGGAGGTGCTGCTGCACTTCCCGAGATGCACTGACGAGATC 834
Qy 825 GAGACCTGCGCTGCGGCAACCACTATTGAGGACCTGACTGACAGCGCTGCGATACG 884
Db 835 GAGACGCTGCGCTGCGGCAACCACTATTGAGGACCTGACTGACAGCGCTGCGATACG 894
Qy 885 GACCAAGCTTCTAGCGCGCGGAGCGCGCTGCTGCG 919
Db 895 GACCACTGCGCGCGCGCGGCTGCTGCGCGCGCG 929

RESULT 11
US-08-552-142A-12
Sequence 12, Application US/08552142A
Patent No. 5695995
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold M.
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/552.142A
FILING DATE: 02-NOV-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas F.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FPCR-1-8933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: 20A1
FEATURE:
NAME/KEY: CDS

LOCATION: 1..310
US-08-552-142A-12
Query Match
Best Local Similarity 65.2%; Score 121.8; DB 1; Length 310;
Matches 191; Conservative 0; Mismatches 101; Indels 1; Gaps 1;

Qy 619 GCCGAGGACATCCAGAGAGCTCCGTGCGCGCGCGAGGCGCAAGGCCAAGAGCG 678
Db 16 GTCAGGGGCAACAG 75
Qy 679 AGTTGGCATGAGGAG 738
Db 76 AGCGGCTGCTGCAACCCCTGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 134
Qy 739 ACCGATGACAACTTAACCTGCGCGCTGATGAGGAGAGAGAGAGAGAGAGAGAG 798
Db 135 ACCGATGACAACTTAACCTGCGCGCGCTGATGAGAGAGAGAGAGAGAGAGAGAG 194
Qy 799 CGGATGAGCGCAACTTAACAAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 858
Db 195 CCGAGAGACCAACTGACCAAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 254
Qy 859 CACTGACTGAGAGCGCTGCGCATAGCGAGACAGAGCTTCTACGCGCCGAGGCC 911
Db 255 CTCTGCGGAGAGACACTGCGCTGCGGATMAAGGCGCTGCGCGAGCGGCGGCC 307

RESULT 12
US-08-722-570-14
Sequence 14, Application US/08722570
Patent No. 655337
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qiduo
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hobbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722.570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-14
Query Match
Best Local Similarity 67.4%; Score 115.6; DB 4; Length 1312;
Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGCGAGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACGACGACGAG 706
DB 539 GAGAGCGCGAGCGCGCGCTCAGGCGCAAGCGGAGAAACTGTGTTAAAGATCAAGAAC 598
QY 707 CCGGCGCAAGAGGCGCAACGAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCGCT 766
DB 599 CCGGCGCGTTAAAGCTAAACAACCGGGAAGAAATCGCATGCAACCTGAACCTCGCGCT 658
QY 767 GGATGCGGTGGCGGTGTCTGCTCCACCTTCCGGATGACGCAAACTTAACAAGATGCA 826
DB 659 TGAATTCCTCAGGAGAGGTGTGCTCTTACCTGAAGATGCAAACTCACAAGATGCA 718
QY 827 GACCTGCGCTTCGCCCAACTACATTTGGGCACTGACGCTGCGCATAGCGGA 886
DB 719 GACCTTGGCTTTGGCTTAACAATCACTGCGGCTCTTAAGCAAACTTGGCGCTTGGCGGA 778
QY 887 CC 888
DB 779 CC 780

RESULT 13
US-08-932-411A-15
; Sequence 15, Application US/08932411A
; Patent No. 6566496

GENERAL INFORMATION:

APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hobach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1312 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-932-411A-15

Query Match 7.9%; Score 115.6; DB 4; Length 1312;
Best Local Similarity 67.4%; Pred. No. 3.1e-21;
Matches 163; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 647 GCGGCGCGAGGCGCAACAGGCCCAAGAGCGAGTTGGCACTGACGACGACGACGAG 706
DB 539 GAGAGCGCGAGCGCGCGCTCAGGCGCAAGCGGAGAAACTGTGTTAAAGATCAAGAAC 598
QY 707 CCGGCGCAAGAGGCGCAACGAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCGCT 766
DB 599 CCGGCGCGTTAAAGCTAAACAACCGGGAAGAAATCGCATGCAACCTGAACCTCGCGCT 658
QY 767 GGATGCGGTGGCGGTGTCTGCTCCACCTTCCGGATGACGCAAACTTAACAAGATGCA 826
DB 659 TGAATTCCTCAGGAGAGGTGTGCTCTTACCTGAAGATGCAAACTCACAAGATGCA 718
QY 827 GACCTGCGCTTCGCCCAACTACATTTGGGCACTGACGCTGCGCATAGCGGA 886
DB 719 GACCTTGGCTTTGGCTTAACAATCACTGCGGCTCTTAAGCAAACTTGGCGCTTGGCGGA 778
QY 887 CC 888
DB 779 CC 780

RESULT 14
US-08-722-570-15
; Sequence 15, Application US/08722570
; Patent No. 6555337

GENERAL INFORMATION:

APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr, Hobach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 1277 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-15

Query Match 7.5%; Score 109.8; DB 4; Length 1277;
Best Local Similarity 74.6%; Pred. No. 1e-19;
Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCGGCGCGAGGCGCAACGAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCG 764
DB 439 ACCCGCGCGTTAAAGCAAAATACCGGAGAGATCGCATGCAACCTGAACCTATGCG 498
QY 765 CTGATGCGCTGCGGCTGTCTCTGCCCACTTCCCGATGACGCCAACTTAACAAGATC 824

Db 499 CTCATTCTGAGGAGGTTCTACCGTCATTACCCGAGCGCAACCTCACCAATA 558
QY 825 GAGACCTGCGCTTGCGCCACCACTACATTGGGCACTGACTGAGCGCTGCGATGCG 884
Db 559 GAGACCTTGCGCTTGCGCCACCACTACATTCTGAGGCTCTTACCGAACTTTGCGCTGCGC 618
QY 885 GACCA 889
Db 619 GACCA 623

Db 559 GAGACCTTGCGCTTGCGCCACCACTACATTGAGGCTCTTACCGAACTTTGCGCTGCGC 618
QY 885 GACCA 889
Db 619 GACCA 623
Search completed: January 26, 2004, 22:06:42
Job time : 100 secs

RESULT 15
US-08-932-411A-16
; Sequence 16, Application US/08932411A
; Patent No. 6566496
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/932,411A
; FILING DATE: 15-SEP-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/772,009
; FILING DATE: 19-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/722,570
; FILING DATE: 19-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-932-411A-16

Query Match 7.5%; Score 109.8; DB 4; Length 1277;
Best Local Similarity 74.6%; Pred. No. 1e-19;
Matches 138; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
QY 705 AGCCGGCGCAGAGCAAGCGGAGCGCAACCGCATGCAACCTTAACCTCGCG 764
Db 439 ACCCGGGCGCTTAAGCAATACCGCGAGAGATCGCATGACACCTGACTATGCG 498
QY 765 CTGATGCGCTGCGCGGTCTCTGCGCACTTCCCGGATGAGCGCAACTTACCAAGTC 824
Db 499 CTGATTTCTGAGGAGGTTCTACCGTCATTACCGGAAAGCCCAACTCACCAAGATA 558
QY 825 GAGACCTTGCGCTTGCGCCACCACTACATTGGGCACTGACTGAGCGCTGCGATGCG 884

THIS PAGE BLANK (USPTO)

Db 714 ACTGCATGACGCTCAGAAATCCCTGCGTCTCATCACTGACGAGTGTGAGTACT 773
Qy 178 CCTCCAGCTCTTGTGCTGCTCCGACGACAACTTACTCCAGGCGGCGCTGACCT 237
Db 774 CTTGGAGCTTTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 833
Qy 238 CAGCAAACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 296
Db 834 TAGCAGAACTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 893
Qy 297 -----ACCCGAGCTCTCTGTTCTTTTGAAGCCCGGAGTA 330
Db 894 TATCCACTGCTGCTTGTCACTGACTGCTGCTGCTTCTTATCTTTTGAAGCTGAGG- 952
Qy 331 ACTAGTAACCTTGAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 380
Db 953 ACTAGTAACCTTGAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 1012
Qy 381 GCGTACTGATCCCGCTGAGTACCTTAACTCAGAGACTG--TCAACCCCTTC 438
Db 1013 GGATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1072
Qy 439 CATTTTTCCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 498
Db 1073 CATTTTTCCTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 1132
Qy 499 CCCAGAGACCCAGAACCTTTCCCGAGCTTCCGAGCCAGAACCTTCAAGCTTCAAGT 558
Db 1133 CCCAGAGACCCAGAACCTTTTCCTGAGCTTCCGAGCCAGAACCTTCAAGCTTCAAGT 1192
Qy 559 CCACCCCACTGAGCTTCTGTAACGAGGAGCTGCTCCGAGCCAGAACCTTCAAGT 618
Db 1193 CCACCCCACTGAGCTTCTGTAACGAGGAGCTGCTCCGAGCCAGAACCTTCAAGT 1252
Qy 619 GCCGAGGACATCGAGAACTTCCGCTGCGGCGCGGAGGCGCAACAGGCTTCAAGG 678
Db 1253 GCCGAGGACATCGAGAACTTCCGCTGCGGCGCGGAGGCGCAACAGGCTTCAAGG 1312
Qy 679 AGTTGGCACTGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 738
Db 1313 AGTTGGCACTGAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 1372
Qy 739 ACCGATGACAACTTAACTCCGCTGAGTGCCTGCGCTGCTGCTGCTGCTGCTGCT 798
Db 1373 ATCGATGACAACTTAACTCCGCTGAGTGCCTGCGCTGCTGCTGCTGCTGCTGCT 1432
Qy 799 CGGATGAGCCAACTTAAAGATGAGACCTTGCCTTCCGCACTTCAATTTGGG 858
Db 1433 CGGATGAGCCAACTTAAAGATGAGACCTTGCCTTCCGCACTTCAATTTGGG 1492
Qy 859 CATGATCTCAAGCTGCGCATAGCGGACCAAGCTTCAAGCTTCAAGCTTCAAGCTT 918
Db 1493 CATGATCTCAAGCTGCGCATAGCGGACCAAGCTTCAAGCTTCAAGCTTCAAGCTT 1552
Qy 919 CTTGTGGGAGCTGGGAGCCCGGAGGCGGAGCTTCCAGCGGAGCTTGGGCTCTATCT 978
Db 1553 CTTGTGGGAGCTGGGAGCCCGGAGGCGGAGCTTCCAGCGGAGCTTGGGCTCTATCT 1612
Qy 979 CCCCACTTCCCAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 1038
Db 1613 CCCCACTTCCCAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTTGAAGCTT 1672
Qy 1039 TCGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1098
Db 1673 TCGAGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1732
Qy 1099 TCGAAGGCGCCAAAGGCGCTGCGGAGGCGCTGCGGAGGAGGAGGAGGAGGAGG 1158
Db 1733 TCGAAGGCGCCAAAGGCGCTGCGGAGGCGCTGCGGAGGAGGAGGAGGAGGAGGAGG 1792
Qy 1159 CTGTCTGAATGAGAGTGAAGGAGCTGAGGAGCTTCCGCTTCTGCTTCTTCAATTA 1218
Db 1793 CTGTCTGAATGAGAGTGAAGGAGCTTCCGCTTCTGCTTCTTCTTCTTCTTCTTCT 1852

Qy 1219 GTCAGTCC 1227
Db 1853 CTTGATTC 1861

RESULT 2
US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

Query Match 46.7%; Score 682.4; DB 14; Length 861;
Best Local Similarity 89.7%; Pred. No. 1.5e-185;
Matches 771; Conservative 0; Mismatches 76; Indels 13; Gaps 3;

Qy 312 TTTCTTGAAGCCCGAGTAACCTTGAAGAACTTCAAGGAGTGAAGAGGGG 371
Db 2 TTTCTTGAAGCTGAGAG-AACTAGTAACATTTGGAAGCTTCAAGGAGTGAAGAGGG 60
Qy 372 AGTGGT-----GGGCGTACTGATGCTGCGGCTGAGTGAAGCTTCAAGTCAAGA 421
Db 61 CGCGGCGGTGATGATGAGGAGTACTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 120
Qy 422 CTG--TCAACCCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 479
Db 121 CTGGCAACACACACCTTCAATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 180
Qy 480 GCGCCACATTCAGAGTGTCCAGAGAGCCAGCAACCTTCCGAGGAGCTTGGAGCCAG 539
Db 181 GCGCTACATTCAGAGTGTCCAGAGAGCCAGCAACCTTCTTCTTCTTCTTCTTCTTCT 240
Qy 540 GAAGTGTCAAGTTCATTTCAACCCCACTGAGCCCACTTCTGTAACGAGGAGTGTCTCC 599
Db 241 GAAGTGTCAAGTTCATTTCAACCCCACTGAGCCCACTTCTGTAACGAGGAGTGTCTCC 300
Qy 600 GAAGCAGAGAGGAGTGTCTGCGAGGAGCATTCAGAGAGCTTCCGCTGCGGCGGAGG 659
Db 301 GAAGCAGAGAGGAGTGTCTGCGAGGAGCATTCAGAGAGCTTCCGCTGCGGCGGAGG 360
Qy 660 GCGAAGGAGGAGGAGGAGTGTGCACTGAGCAAGCAGCAAGCAGGAGGAGGAGGAGG 719
Db 361 GCGAAGGAGGAGGAGGAGTGTGCACTGAGCAAGCAGCAAGCAGGAGGAGGAGGAGG 420
Qy 720 GCGAAGGAGGAGGAGGAGGAGTGTGCACTGAGCAAGCAGCAAGCAGGAGGAGGAGG 779
Db 421 GCGAAGGAGGAGGAGGAGGAGTGTGCACTGAGCAAGCAGCAAGCAGGAGGAGGAGG 480
Qy 780 GGTGTCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 839
Db 481 GGTGTCTGCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 540


```

Db      2775 CTGAACCTTGGGACCAAGAGCCGCTGAGCTCCCAACGGCCCTCGCTGATCGCTCTC 2834
Qy      286  TGACTTGAACCAACCCGACGCTCTCTGTTCTTTTGAACCCGAGTAAGTAACATTTA 345
Db      2835 TATTTCTTTTGGCCCGGTGAGAAAGTAATATTTTGAAGGCTCCGAGAGGAGCGGAGGAA 2894
Qy      346  GGAAC-CTCCAAAGGAGTGAAGAGGAGTGGGAGTGGGCTACTAGTCCGCGTGAAGT 404
Db      2895 AGAGGAGTCTCTGACCCAGCGGGGCTGGAGAGATGGCTGTTTTTTTTTTTCCCACT 2954
Qy      405  GACCTTAAGTCAAGACTG-----TCACACCCCTTCATTTTTCGAA 451
Db      2955 AGCCTCGAGATCGCGACTGCGCCCGTGAACGAGCTCAAACTTACCTTCCCTCGACCCG 3014
Qy      432  CCTGAGATGGGCTCTATCCCTTGGATGGCCCAACATCCAAAGTGTCCCAAGAACCA 511
Db      3015 CCGTGGAGTGAAGCTTCAACCTTCGGGTGGCCCACTGTCCAAAGTACCCCGTGAACGA 3074
Qy      512  GCAACCTTTTCCGGAGCTCGGACCAAGAGTGTCAAGTTCCAAATTCACCCCACTAG 571
Db      3075 GCGGTCTTCTCCCAAGGCTTCGGAAGAGAGTGAAGTACCTCCCACTCCGCCCCAG 3134
Qy      572  CCCCCTCTCTGTAACGAGGACTGTCTCCAGACAGAGAGTGAAGTCCGAGGAGCAATC 631
Db      3135 CCCCCTCGACACAGGGGGGAACTGGCGCAGAGCGGAAAGAGAGGAGGCTGCGAGGGGCCC 3194
Qy      632  GAGGAAGTCCGCTGCGCGCGCGCGGAGGCGCAACAGGCCCAAGAGCGAGTTGGACTGAG 691
Db      3195 GAGGAAGTCCGCGCGCGCGCGGAGGAGCGCGCGCTTAAGACGAGTTGGCACTGAG 3254
Qy      692  CAAGAGGAGGAGAGCGCGCGCGCAAGAGGCAAGCAAGCGGGAGCGCAACCGCATGACAA 751
Db      3255 CAAGAGGAGGAGAGTGGCGCAAGAGGCAAGCAAGCGCGAGCGCAATCGATGACAA 3314
Qy      752  CTTTAACCTCCGCGCTGATGCGCTGCGCGCTGCTCTCCCACTTCCCGAGTGAAGC 811
Db      3315 CCTCAACTCGGACCTGAGAGCGCTGCGCGGTGTCTCGGCCACTTCCAGAGAGCGGAA 3374
Qy      812  ACTTAAGAAAGATGAGAGCCTTCCGCTGCGCCCAAACTAATTGGGCACTGACTGAGC 871
Db      3375 GCTTACCAAGATGAGAGCGCTGCGCTGCGCCCAAACTAATCTGGGCGCTGACTCAAC 3434
Qy      872  GGTGGGATGAGGAGGAGCAAGCTTCAAGGCGCCGAGCGCCCTGAGCGCTGAGGAGT 931
Db      3435 GCTGGGATGAGGAGGAGCAAGCTTGTACCGCTGAGAGCGCGCGCGCACTGCGGAG 3494
Qy      932  GGAAGCGCGGAGGAGGCTTCCAGCGCGCACTGGGCGCTTATCTAATCCCACTTTTCCA 991
Db      3495 GGTGGGAGCGCCAGAGGGGTTCCCGGGGAGCTGGGGGTCCCTCTAATCCCACTTCCA 3554
Qy      992  AGCTGTAGCTGAGAGCCCAAGAGCTCATTTGAGAGATTCCTGGCGCTGAGAGTCCCG 1051
Db      3555 GGTGGGAGCGCTGAGAGTCCCGCGCTGCTGAGAGGAGCAACCGGCGCTGCTGGGCGCAC 3614
Qy      1052 CTCCCATCTCTGCTGCTGCTGCGGAGCAACCTGAGTTCTGAGCTTCTGAAAGGCGCA 1111
Db      3615 CTCTTCGCTGCTGAGAGCTGAGGCTGAGGCTTGTGCTTTCTAGATTTTCTGAAAGGACT 3674
Qy      1112 AACAGGCTTGGGCGGCTGCGCTGAGCAAGAGGAGGAGTCAAGCTGTCTGAATAG 1171
Db      3675 GTCTGTGCTGGGCTGTGGGTCTAAGGTAAGGAGAGGAGGAGCGGGAGCGCTAG 3734
Qy      1172 AAGGTAGTGAAGCACTGAGCAATC 1196
Db      3735 AGGTGGCGGAGCGGCGGCGCTC 3759

```

RESULT 5
 US-08-722-570-12
 ; Sequence 12, Application US/08722570
 ; Publication No. US20030044887A1
 ; GENERAL INFORMATION:

```

; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
;
; TELEX: 910 277299
;
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1527 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
;
; US-08-722-570-12

```

Query Match 9.9%; Score 145.2; DB 8; Length 1527;
 Best Local Similarity 69.2%; Pred. No. 2e-31; 88; Indels 0; Gaps 0;
 Matches 198; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

```

Qy      635  GAAAGCTCCGTGCGCGCGCGGAGGCGGCAAGAGCGCAAGAGTGGCACTGAGCA 694
Db      456  GAAGAGGCGGCGGCGGAGGCGGAGGCGGCGGAGTGGCGGCGGCGGCGGCGGCGG 515
Qy      695  GCAAGGAGGAGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 754
Db      516  GCTGGGAGGAGGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 575
Qy      755  TAACTCCGCGCTGATGCGCTGCGCGGCTGCTGCTGCGGCACTTCCCGAGTGAAGCCAACT 814
Db      576  CAAGCTGCGCTGAGAGCGCTGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAG 635
Qy      815  TACAAAGATGAGAGCCTGCGGCTTGCACCAACTAATTGGGAGCACTGACTCAGAGCT 874
Db      636  CACCAAGATGAGAGCGCTGCGGCTTGCCTTACCAACTAATCTGAGGCGCTGAGAGACT 695
Qy      875  GCGCATAGCGGAGCACACACTTCTAGCGGCCCGGAGGCCCTGTGCCC 920
Db      696  GCGCTGCGAGTCAAGGCTTCCCGGGGAGGCTGCGGAGGCGG 741

```

RESULT 6
 US-08-722-570-13
 ; Sequence 13, Application US/08722570
 ; Publication No. US20030044887A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Anderson, David J.
 ; APPLICANT: Ma, Qifu
 ; TITLE OF INVENTION: NEUROGENIN
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Flehr, Hohbach, Test, Albritton & Herbert

STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/08/722,570
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-13

Query Match 9.6%; Score 140.4; DB 8; Length 738;
Best Local Similarity 68.2%; Pred. No. 4,1e-30;
Matches 195; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 635 GAAGCTCCGTCGCGCGCGGAGGCGGCAAGCGCCCAAGAGGATTGGACCTGAGCAA 694
DB 207 GAGAGAGCGGCGGAGCGCGGAGGTCGCTGGGTGGCGTCCAGGCTGCTGCTCACTC 266
QY 695 GCGAGCGAGAGCGCGCGGCGGAGGCGGCAAGCGGAGGCGGCAAGCGGAGGCGG 754
DB 267 CCTGGGAGGAGTCTCGCTGCAAGCGGAGGCGGAGGCGGCGGAGGCGGAGGCGG 326
QY 755 TAACCTCGCGCTGATGCGGTCGCGGCGGTCCTGCTGCTGCTGCTGCTGCTGCTGCT 814
DB 327 CAAGCTCGCTGATGCGGTCGCGGCGGTCGCTGCTGCTGCTGCTGCTGCTGCTGCT 386
QY 815 TACAAAGATCGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 874
DB 387 CACCAAGATTGAGAGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 446
QY 875 GCGCATGCGGAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 920
DB 447 GCGCTGCGAGATCAAGGAGCTGCCGCGGAGGAGGCGGAGGCGGAGGCGGAGGCGG 492

RESULT 7

US-10-004-717-30
Sequence 30, Application US/10004717
Publication No. US2002019265A1
GENERAL INFORMATION:
APPLICANT: ZOGHBI, HUDA Y.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
FILE REFERENCE: P01899US4
CURRENT FILING DATE: 2002-08-16
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: 09/585,645
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: 60/176,993
PRIOR FILING DATE: 2000-01-19

PRIOR APPLICATION NUMBER: 60/137,060
PRIOR FILING DATE: 1999-06-01
NUMBER OF SEQ ID NOS: 69
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 30
LENGTH: 1385
TYPE: DNA
ORGANISM: Mus musculus
US-10-004-717-30

Query Match 9.6%; Score 139.8; DB 14; Length 1385;
Best Local Similarity 78.1%; Pred. No. 7e-30;
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGGCGCAAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 764
DB 715 ACCCGAGCTCAAGGCGCAAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 774
QY 765 CTGATGCGCTGCGCGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
DB 775 CTGAGCGGCTGCGCGGAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 834
QY 825 GAGACCTGCGCTGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 884
DB 835 GAGACGCTGCGCTGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 894
QY 885 GACCAGAGCTTCTACGGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 919
DB 895 GACCACTGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 929

RESULT 8

US-10-004-717-6
Sequence 6, Application US/10004717
Publication No. US2002019265A1
GENERAL INFORMATION:
APPLICANT: ZOGHBI, HUDA Y.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
FILE REFERENCE: P01899US4
CURRENT FILING DATE: 2002-08-16
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: 09/585,645
PRIOR FILING DATE: 2000-06-01
PRIOR APPLICATION NUMBER: 60/176,993
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: 60/137,060
PRIOR FILING DATE: 1999-06-01
NUMBER OF SEQ ID NOS: 69
SOFTWARE: Patent In Ver. 2.1
SEQ ID NO 6
LENGTH: 1412
TYPE: DNA
ORGANISM: Mus musculus
US-10-004-717-6

Query Match 9.6%; Score 139.8; DB 14; Length 1412;
Best Local Similarity 78.1%; Pred. No. 7e-30;
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGGCGCAAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 764
DB 418 ACCCGAGCTCAAGGCGCAAGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 477
QY 765 CTGATGCGCTGCGCGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
DB 478 CTGAGCGGCTGCGCGGAGGTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 537
QY 825 GAGACCTGCGCTGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 884
DB 538 GAGACGCTGCGCTGCGGCGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGGCGGAGG 597

QY 885 GACCAAGCTTCTACGCGCCGAGCCCTCTGTGCC 919
DB 598 GACCACTGCGCGCGCGGTGCTCCACGAGGGGC 632

RESULT 9

US-10-004-717-37
; Sequence 37, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-37

Query Match 9.6%; Score 139.8; DB 14; Length 1412;
Best Local Similarity 78.1%; Pred. No. 7e-30;
Matches 168; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 705 AGCCGCGCAAGAGCCCAACGAGCGGAGCCCAACCGCATGCACAACTTAATCCGGC 764
DB 418 ACCCGAGGCTCAAGGCCCAACACCGGAGCCGATGCACAACTTAACGCGCG 477
QY 765 CTGATGCGCTGCGCGGTCTCTGCGCCACTTCCCGATGAGCGCAACTTAACAAGATC 824
DB 478 CTGAGCGCTGCGCGAGGTGCTGCGCCACTTCCCGAGATGCCAAGCTCAGAAATC 537
QY 825 GAGACCGCGCTGCGCGCCCAACTGATGCGCTGATGCGCTGCGCTGCGCTGCGCT 884
DB 538 GAGACGCTGCGCTGCGCCCAACTGATGCGCTGATGCGCTGCGCTGCGCTGCGCT 597
QY 885 GACCAAGCTTCTACGCGCCGAGCCCTCTGTGCC 919
DB 598 GACCACTGCGCGCGCGGTGCTCCACGAGGGGC 632

RESULT 10

US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01

; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-20

Query Match 9.5%; Score 138.2; DB 14; Length 790;
Best Local Similarity 84.8%; Pred. No. 9.3e-30;
Matches 156; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 702 CGAAGCCGCGCAAGAGCCCAACGAGCGGAGCCCAACCGCATGCACAACTTAATCC 761
DB 344 CGGAGCCGCGCGCTGAAAGCCCAACCGCGCAACCGCATGCACAACTTAATCC 403
QY 762 GCGCTGATGCGCGCGGTCTCTGCGCCACTTCCCGATGAGCGCAACTTAACAAG 821
DB 404 GCGCTGAGCGCGCTGCGGAGGTGCTGCGCCACTTCCCGAGAGCCCAAGCTCACCAAG 463
QY 822 ATCGAGCCCTGCGCTGCGCCCAACTGATGCGCTGATGCGCTGCGCTGCGCTGCGCT 881
DB 464 ATCGAGCGCTGCGCTGCGCCCAACTGATGCGCTGATGCGCTGCGCTGCGCTGCGCT 523
QY 882 GCGG 885
DB 524 GCGG 527

RESULT 11

US-10-004-717-18
; Sequence 18, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1074
; TYPE: DNA
; ORGANISM: chicken
US-10-004-717-18

Query Match 9.5%; Score 138.2; DB 14; Length 1074;
Best Local Similarity 71.4%; Pred. No. 1.9e-29;
Matches 182; Conservative 0; Mismatches 73; Indels 0; Gaps 0;

QY 642 CGTGGCGCGCGCGAGGCGCAACGCGCCCAAGAGCGATGCGCTGAGCAAGAGCGA 701
DB 372 CGCGGAGAGAGCGCGCGGTGCGGCGCGCGAGCGGAGCTTGTGCGACACCTTCAAA 431
QY 702 CGAAGCCGCGCAAGAGCCCAACGAGCGGAGCCCAACCGCATGCACAACTTAATCC 761
DB 432 CGGAGCCGCGCGGTAAAGCCCAACGCGGAGCGGAGCCGATGCGCTGAGCAAGCGC 491
QY 762 GCGCTGATGCGCGCGGTCTCTGCGCCACTTCCCGATGAGCGCAACTTAACAAG 821
DB 492 GCGCTGATGAGCTCGAGCGCTCTGCGGAGCTTCCCGAGAGCAACCACTTCAACAA 551
QY 822 ATCGAGCCCTGCGCTGCGCCCAACTGATGCGCTGATGCGCTGCGCTGCGCTGCGCT 881

Db 552 ATCGAAACCTGCTGCTTACCAATCGAGCCCTCTCGAAGACCTTCGTTTG 611
QY 882 GCGAGCACAGCTTC 896
Db 612 GCCGAGCAGTGCTTC 626

RESULT 12

US-08-722-570-14
; Sequence 14, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-722-570-14

Query Match 7.9%; Score 115.6; DB 8; Length 1312;
Best Local Similarity 67.4%; Pred. No. 6e-23; Indels 0; Gaps 0;
Matches 163; Conservative 0; Mismatches 79;

QY 647 GCGGCGGAGGCGCAACAGGCGAGTGGCACTGACAGCAGCAGCAAG 706
Db 539 GAGGAGCCGAGGCGGCTCAGGCGAAGAGCGGAGAACTGTGTTAAAGATCAAGAAAC 598
QY 707 CCGGCGCAAGAGGCGCAACGACCGGAGCGCAACCGCATGCAACCTTAACTCGGCGCT 766
Db 599 CCGGCGCGTTAAAGCTAACACCGGGAAGAAATGCGATGCAACCTGAACCTCGGCGCT 658
QY 767 GGATGCGGTGGCGGTCTCTGCGCACTTCCCGATGACGCGCAACTTCAAGATGCA 826
Db 659 TGATTCCTCAAGGAGGTGTGCTCTTACTTAAGATGCAAACTCACCAAGATAGA 718
QY 827 GACCTGCGCTTGGCCCAACTATATTGGGCACTGACTGACGCTGGCATAGCGGA 886
Db 719 GACCTTGGCTTTGGCTTACACTATCTGGGCTCTTAGCGAAACTTGGCGCTTGGCGGA 778
QY 887 CC 888
Db 779 CC 780

RESULT 13

US-08-722-570-15
; Sequence 15, Application US/08722570
; Publication No. US20030044887A1
; GENERAL INFORMATION:
; APPLICANT: Anderson, David J.
; APPLICANT: Ma, Qifu
; TITLE OF INVENTION: NEUROGENIN
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/722,570
; FILING DATE: 27-SEP-1996
; CLASSIFICATION: 5365
; ATTORNEY/AGENT INFORMATION:
; NAME: Silva, Robin M.
; REGISTRATION NUMBER: 38,304
; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; US-08-722-570-15

Query Match 7.5%; Score 109.8; DB 8; Length 1277;
Best Local Similarity 74.6%; Pred. No. 2.8e-21; Indels 0; Gaps 0;
Matches 138; Conservative 0; Mismatches 47;

QY 705 AGCCGCGCAAGAGGCGCAACGCGGAGCGCAACCGCATGCAACCTTAACTCGCGG 764
Db 439 ACCCGGCGCGTTAAAGCCAAATACCGGAGAGATGCGATGCAACCTGAATATGCG 498
QY 765 CTGATGCGCTGCGCGGTGCTCTGCCCACTTCCCGATGAGCGCAACCTTAAAGATC 824
Db 499 CTCGATTCCTGAGGAGGAGTCTACCGTCAATTACCGAAGCGCAAACTCACCAAGATA 558
QY 825 GAGACCCCTGCGCTTCCGCGCAACTATATTGGGCACTGACTGAGCGCTGGCATGGC 884
Db 559 GAGACCTTGGCTTGGCCCAACTATATCTGGGCTCTTAGCGAACTTGGCGCTGGCC 618
QY 885 GACCA 889
Db 619 GACCA 623

RESULT 14

US-10-413-358-27
; Sequence 27, Application US/10413358
; Publication No. US20030219894A1
; GENERAL INFORMATION:
; APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.
; TITLE OF INVENTION: Induction of insulin-producing cells
; FILE REFERENCE: GP55

Job time : 520 secs

; CURRENT APPLICATION NUMBER: US/10/413,358
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: JP2002-115064
; PRIOR FILING DATE: 2002-04-17
; NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 27
; LENGTH: 1099
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-413-358-27

Query Match 6.1%; Score 89; DB 13; Length 1099;
Best Local Similarity 61.4%; Pred. No. 2.5e-15;
Matches 143; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 650 GCGCGAGAGCGCCACAGCCCAAGAGAGTGGCACTGAGCAGCAGCAGAGCCG 709
DB 258 GAGACGCGCGCCCAAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 317
QY 710 GCGCAAGAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769
DB 318 AGCATGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 377
QY 770 TGGCGTGGCGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 829
DB 378 CAACCTGGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 437
QY 830 CTTGGCGCTTGGCGCAATTAATTTGGGCACTGAGAGCTGCGCATAG 882
DB 438 TCTGGCGCTTGGCGCAATTAATTTGGGCACTGAGAGCTGCGCATAG 490

RESULT 15

US-10-413-358-26
; Sequence 26, Application US/10413358
; Publication No. US20030219894A1
; GENERAL INFORMATION:
; APPLICANT: Susumu Seino, JCR Pharmaceuticals Co., Ltd.
; TITLE OF INVENTION: Induction of insulin-producing cells
; FILE REFERENCE: GPs5
; CURRENT APPLICATION NUMBER: US/10/413,358
; CURRENT FILING DATE: 2003-04-15
; PRIOR APPLICATION NUMBER: JP2002-115064
; PRIOR FILING DATE: 2002-04-17
; NUMBER OF SEQ ID NOS: 27
; SEQ ID NO 26
; LENGTH: 1211
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-413-358-26

Query Match 6.1%; Score 89; DB 13; Length 1211;
Best Local Similarity 61.4%; Pred. No. 2.5e-15;
Matches 143; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 650 GCGCGAGAGCGCCACAGCCCAAGAGAGTGGCACTGAGCAGCAGCAGAGCCG 709
DB 348 GAGACGCGCGCCCAAGAGAGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 407
QY 710 GCGCAAGAGAGCGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 769
DB 408 AGCATGAAGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 467
QY 770 TGGCGTGGCGGAGTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 829
DB 468 CAACCTGGCGAGAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 527
QY 830 CTTGGCGCTTGGCGCAATTAATTTGGGCACTGAGAGCTGCGCATAG 882
DB 528 TCTGGCGCTTGGCGCAATTAATTTGGGCACTGAGAGCTGCGCATAG 580

Search completed: January 26, 2004, 22:15:27

Db		388	GAACTGGGTGACTGCTCCGAGGAGACTTCGAGGAAAGCTCCGCGCCGACGCGGAGGGCGCAAC	447
QY		666	AGGCCCAAGAGCGAGTTGGGCACTGAGCAAGACGACGAAAGCGGCGGCAAGAGGCCAAC	725
Db		448	AGGCCCAAGAGCGAGTTGGGCACTGAGCAAAAGCGAAGAGCGCGCAAGAGGCCAAT	507
QY		726	GACCGGGAGCGGCACACCGCATGACAACTTAACTCCGCGCTGAGATGCCCTGGCGCGGTGC	785
Db		508	GATCGGAGCGCAATCGCATGACAACTTCAACTGCGCGCTGAGATGCCCTGGCGCGGTGC	567
QY		786	CTGCCACACTTCCCGGATGAGCGCCAACTTAAAGATGAGAACCTTGCGCTTCCGCCAC	845
Db		568	CTGCCACACTTCCCGGATGAGCGCCAACTTAAAGATGAGAACCTTGCGCTTCCGCCAC	627
QY		846	AACTACATTTGGGCACTGACTGACGCTGCGCATAGCGGACACACAGCTTCTACGSCCCC	905
Db		628	AACTACATCTGGGCACTGACTGACGCTGCGCATAGCGGACACACAGCTTGTATGGCCCG	687
QY		906	GAGCCCCCTTGGCTCTGTGGGAGAGTGGGGAAGCCCGGAGAGGGGGCTTCCAGCGGCGCATCG	965
Db		688	GAGCCCCCTTGGCTCTGTGGGAGAGTGGGGAAGCCCGGAGAGGGGGCTTCCAGCGGCGCATCG	747
QY		966	GAGCTTATCTACTCCCAAGTTTCCCAAGGTGGTAGCTGAGCGCCACAGCCCTCATTTGGAG	1022
Db		748	GAGCTTATCTACTCCCAAGTTTCCCAAGGTGGTAGCTGAGCGCCACAGCCCTCATTTGGAG	807
QY		1026	GAGTTCCCTGGGCTGACAGGTGCCACGCTCCCATCTGTCTGCTCCGCGGACCCCTGGTG	1085
Db		808	GAAATTCCTGGGCTGACAGGTGCCACGCTCCCATCTGTCTGCTCCGCGGACACCTGGTG	867
QY		1086	TTCTCAGACTTCTTGTGAAGGGGCCAAACAGGCGCTTGGGCGGTGGGCGCTGGCAGAAAGG	1145
Db		868	TTCTCAGACTTCTTGTGAAGGGGCCAAACAGGCGCTTGGGCGGTGGGCGCTGGTGAAGAGG	927
QY		1146	GAGGAGATCAGAGCGTGTGAAATGGAAAGTATGTGAGGCACTGACAGATCTCGGCCCTT	1200
Db		928	GAGGAGATCAGAGCGTGTGAAATGGAAAGTATGTGAGGCACTGACAGATCTCGGCCCTT	987
QY		1206	CTGGCTTTCATTAGTCAGAGTCCCT-----GATTTAAACAGAGATTCGACAGT	1252
Db		988	CTGGCTTTCATTAGTCAGAGTCCCT-----GATTTAAACAGAGATTCGACAGT	104
QY		1253	TCCTTGTGCTGTGCTGTCGTCACAAAGATATTTGACAGGCTGATCTCTTTAAACCTCTCA	1312
Db		1048	TAC---CCCTGTGTGGCCACAGACAGACATTTGGGGGCTGTCTTCTTAACTCTCTCG	1104
QY		1313	GTTGGGCAACCTCAAACTCCGCTCCAGACAGAGAACGCTTACCACTTAATAGTTGGG	1372
Db		1105	GTTGGGCAACCTCAAACTCCGCTCCAGACAGAGAACGCTTACCACTTAATAGTTAGT--GG	1166
QY		1373	AGACTCCCATACTTCTCTGTGACTCCGCTCTTTCAATCTGCGGGCCT-----CCAAC	1428
Db		1164	AGACTCCCATACTTCTCTGTGACTCCGCTCTTTCAATCTGCGGGCCTCCGACCATTC	1222
QY		1429	ACCGCTTCTCCAGAGTACCTTAATCCAGTGT	1460
Db		1224	ATCACTTTTCCAGGGTACCTTAATCCAGTGT	1255
RESULT 2				
BY708009				
LOCUS		1025 bp	mRNA	linear
DEFINITION		BY708009 RIKEN full-length enriched, adult male small intestine Mus		EST 16-DEC-2000
ACCESSION		BY708009		
VERSION		BY708009.1		GI:27119192
KEYWORDS				
SOURCE		Mus musculus (house mouse)		
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
		Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
REFERENCE		1 (bases 1 to 1025)		

AUTHORS
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
 Nikaide, I., Osato, N., Saito, R., Suzuki, H., Yamanaoka, I., Kiyosawa, H.,
 Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
 Gotohori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A.,
 Quackenbush, J., Schriml, L. M., Kapin, A., Matsuda, H., Batalov, S.,
 Batseil, K. W., Blake, J. A., Brad, D., Brusic, V., Chochocka, C., Corbett
 L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest
 A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A.,
 Gough, J., Grimmond, S., Gustlich, S., Hirokawa, S., Jackson, I. J.,
 Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R. M.,
 King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons
 P. A., Maglott, D. R., Maltais, L., Marchionni, L., McKenzie, L., Mikki
 H., Nagashima, T., Numata, K., Okido, T., Pavani, W. J., Pereira, G.,
 Pesole, G., Petrovsky, N., Pillai, R., Ponting, J. U., Qi, D.,
 Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring
 B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou
 M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale
 R. D., Tomita, M., Verardo, R., Wagner, L., Wahlstedt, C., Wang, Y.,
 Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yangisawa
 M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
 Carninci, P., Hayatsu, N., Hirotsune-Kishikawa, T., Komno, H., Nakamura
 M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
 Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii
 Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata
 K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander
 E. S., Rogers, J., Birney, E. and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 22354683
 12466851
 COMMENT
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 FEATURES
 SOURCE
 Location/Qualifiers
 1. 1025
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="2010001M19"
 /sex="male"
 FEATURES
 SOURCE
 Location/Qualifiers
 1. 1025
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /clone="2010001M19"
 /sex="male"

ECORI; Site 2: EcorI; Female C57BL/6J mouse kidney and/or brain genomic DNA was isolated and partially digested with a combination of EcorI and EcorI Methylase. Site selected DNA was cloned into the pBAC3.6 vector at the EcorI sites. The ligation products were transformed into DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 88 a 159 c 213 g 133 t
ORIGIN

Query Match 32.4%; Score 473.4; DB 28; Length 593;
Best Local Similarity 91.5%; Pred. No. 1.8e-105;
Matches 535; Conservative 0; Mismatches 46; Indels 4; Gaps 3;

426 CACACCCCTTCCATTTTTCACCACTCAGATGAGCGCTCATCTTGATGCGCC 485
586 CACACACCTTCCATTTTTCACCACTCAGATGAGCGCTCATCTTGATGCGCTC 527
486 ACCATCCAGTGTCCCAAGACCAAGCAACCTTTCCCGAGCTCGAGCCAGAGTG 545
526 ACCATCCAGTGTCCCGAGACCAAGCAACCTTTCCCGAGCTCGAGCCAGAGTG 467
546 CTCAGTCCATTCACCCCACTCAGCGCCCTCTCGTACGAGGAGCTGCTCGAGCA 605
466 CTCAGTCCATTCACCCCACTCAGCGCCCTCTCGTACGAGGAGCTGCTCGAGCA 407
606 GAAGCAGGTGACTGCGGAGGACATCGAGAGAGCTCCGTCGCGCGCGAGGCGCAAC 665
406 GAAGTGGTGAATCGCGAGGACCTCGAGAGAGCTCCGTCGCGCGAGGCGCGCAAC 347
666 AGGCCCAAGAGGAGTGTGCACTGAGCAAGAGCAGCAAGCGCGCGCAAGAGCGCAAC 725
346 AGGCCCAAGAGGAGTGTGCACTGAGCAAGAGCAGCAAGCGCGCGCAAGAGCGCAAT 287
726 GACCGGAGCGCAACCGATGCAACCTTAACTCCGCGTGGATGCGCTCGCGGTC 785
286 GATCGGAGCGCAACCGATGCAACCTTAACTCCGCGTGGATGCGCTCGCGGTC 227
786 CTGCGCACCTTCCCGAGTACGCGCAAACTTACAAAGATGAGAGCCCTGCTCGCCAC 845
226 CTGCGCACCTTCCCGAGTACGCGCAAACTTACAAAGATGAGAGCCCTGCTCGCCAC 167
846 AACTCATTTGGGCACTGACTGAGAGCTGCGCATAGCGGACCAAGCTTTCAGGCGCC 905
166 AACTCATTTGGGCACTGACTGAGAGCTGCGCATAGCGGACCAAGCTTTCAGGCGCC 107
906 GAGCCCTGTCGCTGCGGAGAGCTGCGGAGCGCGGAGCGGCTTCCAGCGCGCACTCG 965
106 GAGCCCTGTCGCTGCGGAGAGCTGCGGAGCGCGGAGCGGCTTCCAGCGCGCACTCG 49
966 GAGCTTATCTACTCCCGAGTTTCCCAAGCTGTAGCTGAGCGCC 1010
48 GAGCTT-ATCACTCCCGAGTCTCCCAAGCGGAGCACTTACGCGCCAC 6

RESULT 4
CA945402 687 bp mRNA linear EST 30-DEC-2002
LOCUS
DEFINITION
IMAGE: 6828925 5', mRNA sequence.
UT-M-FD0-cdh-1-12-0-UT.r1 NIH_BMAP_F00 Mus musculus cDNA clone

ACCESSION
CA945402
VERSION
CA945402.1 GI:27433882
KEYWORDS
EST.
SOURCE
Mus musculus (house mouse)

ORGANISM
Mus musculus (house mouse)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 687)
NIH-MGC <http://mgi.mc.ncl.nih.gov/>
National Institutes of Health, Mammalian Gene Collection (MGC)

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa

cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA library arrayed by: Dr. M. Bento Soares, University of Iowa
DNA sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
<http://image.lnl.gov>

This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

FEATURES
source

Seq primer: pYX-5.

Location/Qualifiers

1.687
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6828925"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_id="NIH BMAP FD0"

/note="Organ: brain; Vector: pYX-Abs; Site 1: Ecor I;
Site 2: Not I; The library was constructed according to
Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
1996. Denatured mRNA was size fractionated on a 1% agarose
gel. First strand cDNA synthesis was primed with an
oligo-dT primer containing a Not I site. Double stranded
cDNA was size selected according to mRNA size fraction,
ligated with Ecor I adaptor, digested with Not I, and then
cloned directionally into pYX-Abs vector. The library tag
sequence located between the Not I site and the polyA tail
is TGAGAGAGCC. This library was created for the
University of Iowa Mouse Brain Molecular Anatomy Project
(BMAP): 'Gene Discovery in the Developing Mouse Nervous
System', supported by National Institutes of Mental Health
(NIMH), Hemlin Chin, Ph.D., program coordinator."

BASE COUNT 117 a 228 c 250 g 91 t 1 others
ORIGIN

Query Match 9.1%; Score 133.4; DB 14; Length 687;
Best Local Similarity 83.1%; Pred. No. 5.1e-22;
Matches 152; Conservative 0; Mismatches 31; Indels 0; Gaps 0;

705 AGCCGGCGCAAGAGGCAAGAGCGGAGGCGCAACCGCATGCAACCTTAATCGCG 764
422 ACCCGAGGCTCAAGGCAAGAGCGGAGGCGCAACCGCATGCAACCTTAATCGCG 481
765 CTGATGCGCTGCGCGGTGCTGCTCCCACTTCCCGATGAGCGCAAACTTAAGATC 824
482 CTGAGCGGCTGCGCGGTGCTGCTCCCACTTCCCGATGAGCGCAAACTTAAGATC 541
825 GAGACCTGCGCTGCGCGCAACTTATGAGCTGAGAGCGCGGATAGCG 884
542 GAGACCTGCGCTGCGCGCAACTTATGAGCTGAGAGCGCGGATAGCG 601
885 GAC 887
602 GAC 604

RESULT 5
AL540071

LOCUS
DEFINITION
IMAGE: 1001 bp mRNA linear EST 31-MAY-2003
UT-A540071 Homo sapiens FETTL BRAIN Homo sapiens cDNA clone

ACCESSION
AL540071
VERSION
AL540071.2 GI:31264632
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 1001)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.

Query Match 8.8%; Score 128.6; DB 12; Length 600;
 Best Local Similarity 79.6%; Pred. No. 7.3e-21;
 Matches 152; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 729 CGGAGCGCCACCGGATGACAACTTAACTCCGCGCTGGAGTCCGCTCGGCTGCTG 788
 Db 1 CGCGAGCGCCACCGGATGACAACTTAACTCCGCGCTGGAGTCCGCTCGGAGTGTG 60
 QY 789 CCCACCTTCCGCGAGTGGAGCCCAACTTAAAGATGAGACCCCTGCGCTGCGCCACAAC 848
 Db 61 CCCACCTTCCGCGAGTGGAGCCCAACTTAAAGATGAGACCCCTGCGCTGCGCCACAAT 120
 QY 849 TACATTTGGGCACTGACTGAGACGCTGGCGATAGCGGACCAAGCTTCTACGCCCCGAG 908
 Db 121 TACATCTGGGGGCTCACCGAGACTCTGGGCGCTGGGAGCACTGGCGCGCGCGGTGGC 180
 QY 909 CCCCTGTGGCC 919
 Db 181 CTCGAGGGGGC 191

RESULT 8
 BUE12495 730 bp mRNA linear EST 20-FEB-2003
 LOCUS UI-M-FR0-cbc-k-21-0-UI.r1 NIH BMAP FR0 Mus musculus cDNA clone

ACCESSION BUE12495
 DEFINITION UI-M-FR0-cbc-k-21-0-UI 5', mRNA sequence.

KEYWORDS BUE12495.1 GI:23278710
 SOURCE EST.
 ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 730)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@uiowa.edu
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES
 SOURCE
 Location/Qualifiers
 1..730
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-FR0-cbc-k-21-0-UI"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /note="Organ: Brain; Vector: pYX-Asc; Site: 1; Ecov I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACG. This library was created for the University Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National

BASE COUNT 131 a 244 c 264 g 91 t
 ORIGIN

Query Match 8.8%; Score 128.6; DB 13; Length 730;
 Best Local Similarity 81.4%; Pred. No. 7.9e-21;
 Matches 149; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

QY 705 AGCGGCGCGCAAGAGGCGCAAGCGGAGGCGCAACCGCATGACCAACTTAACCTCGCG 764
 Db 543 ACCGCGAGCTCAAGGCGCAACCGGAGGCGCAACCGCATGACCAACTTAAGCGCGC 602
 QY 765 CTGATGGCGCTCGCGGTGTCTCTCCCACTTCCCGATGACGCGCAACTTAACAAGATC 824
 Db 603 CTGAGCGGCTCGCGGAGGTCTCTCCCACTTCCCGAGATGCAAGCTACGAGATC 662
 QY 825 GAGACCTGGCGCTTCCCGCACCACTATTGGGCACTGACTCAGACGCTCGCATACG 884
 Db 663 GAGACGCTGGCGCTTCCCGCACCACTATTGGGCACTGACTCAGACGCTCGCATACG 722
 QY 885 GAC 887
 Db 723 GAC 725

RESULT 9
 BUE54481 823 bp mRNA linear EST 26-AUG-2002
 LOCUS UI-M-FD0-bzj-i-24-0-UI.r1 NIH BMAP FD0 Mus musculus cDNA clone

ACCESSION BUE54481
 DEFINITION UI-M-FD0-bzj-i-24-0-UI 5', mRNA sequence.

KEYWORDS BUE54481.1 GI:22494558
 SOURCE EST.
 ORGANISM Mus musculus (house mouse)

REFERENCE 1 (bases 1 to 823)
 AUTHORS NIH-MGC http://mgc.nci.nih.gov/
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 http://image.lnl.gov
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES
 SOURCE
 Location/Qualifiers
 1..823
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6404447"
 /dev_stage="embryo 12.5 dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /note="Organ: brain; Vector: pYX-Asc; Site: 1; Ecov I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then

cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT 155 a 273 c 289 g 104 t 2 others

Query Match 8.7%; Score 127.4; DB 13; Length 823;

Best Local Similarity 81.6%; Pred. No. 1.6e-20;

Matches 146; Conservative 0; Mismatches 33; Indels 0; Gaps 0;

QY 705 AGCCGCGCCAGAAAGCCCAAGCAGCGGAGCGCAACCGATGACCAACTTAACCTCCGG 764
 DB 645 ACCCGAGGCTCAAGGCCCAACCGCAGCGCAACCGATGACCAACTTAACCGCGG 704
 QY 765 CTGATGCGCTGCGCGGTCTCTGCCCACTTCCCGATGACCGCAACTTAAGATC 824
 DB 705 CTGAGCGGCTGCGGAGGTGTGCTGCCCACTTCCCGAGATGCCANGCTACGANGATC 764
 QY 825 GAGACCTGCGCTTCCGCCCAACTTATTTGGGCACTGACTCAGACCGTGGCATACG 883
 DB 765 GAGACGCTGCGCTTCCGCCCAACTTATTTGGGCACTGACTCAGACCGTGGCATACG 823

RESULT 10

EX419330 947 bp mRNA linear EST 15-MAY-2003

LOCUS BX419330 Homo sapiens FETAL BRAIN Homo sapiens CDNA clone

DEFINITION CSODP015Y006 5-PRIME, mRNA sequence.

ACCESSION BX419330

VERSION BX419330.1 GI:30765873

KEYWORDS EST.

ORGANISM Homo sapiens (human)

SOURCE

REFERENCE BP 191 91006 EVRY cedex - France

AUTHORS Email: segre@genoscope.cns.fr; Web: www.genoscope.cns.fr

TITLE Library was constructed by life technologies, a division of

JOURNAL Invitrogen. This sequence belongs to sequence cluster 2626.x

COMMENT Contact: Feng Liang Email: fliang@life.techn.com URL: http://fulllength.invitrogen.com/Invitrogen Corporation 1600

Paradey Avenue Genoscope sequence ID: CSODP015DB03QPL.

FEATURES location/Qualifiers

1..947

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="CSODP015Y006"

/tissue_type="FETAL BRAIN"

/dev_stage="fetal"

/clone_lib="Homo sapiens FETAL BRAIN"

/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA

was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-strand cDNA was digested with Not I and

cloned into the Not I and EcoRV sites of the pCMVSPORT 6

vector. Library was not normalized."

BASE COUNT 73 a 393 c 315 g 130 t 36 others

ORIGIN

Query Match 8.3%; Score 121.8; DB 13; Length 947;

Best Local Similarity 72.7%; Pred. No. 4.1e-19;

Matches 152; Conservative 12; Mismatches 44; Indels 1; Gaps 1;

5

QY 704 AAGCGGCGCAAGAGCCCAAGCAGCGGAGCGCAACCGATGACCAACTTAACCTCCGC 763

DB 717 RACCGTGAAGTCAAGGCCCAAGAGCCGAGCAACCGATGACCAACTTAACCGCGGC 776

QY 764 GCTGATGCGCTGCGGCGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTAAGAT 823

DB 777 ACTGAGACCGCTGCGGCGGAGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTAAGAT 836

QY 824 CGAGACCTGCGCTTCCGCCCAACTTATTTGGGCACTGACTCAGACCGTGGCATACG 883

DB 837 MGAGMCTGGGCTTCCG-CGMAAAMAAAMMTGGGCACTACCGARACCTGCGCTGGC 895

QY 884 GAGACCAAGCTTCAAGCGCCCGAGCCG 912

DB 896 GATCACTGCGGCGGCGCGCGCGGCGGCGC 924

RESULT 11

B0924937 600 bp mRNA linear EST 30-OCT-2002

LOCUS B0924937 7103-91 Mouse E14.5 retina lambda ZAP II Library Mus musculus cDNA,

DEFINITION mRNA sequence.

ACCESSION B0924937

VERSION B0924937.1 GI:24428820

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

AUTHORS Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

TITLE 1 (bases 1 to 600)

JOURNAL Mu, X., Zhao, S., Pershad, R., Hejeh, T., F., Scarpa, A., Wang, S., W.,

MEDLINE White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.

PUBMED Gene expression in the developing mouse retina by EST sequencing

and microarray analysis

CONTACT: Klein WH

DEPARTMENT: Department of Biochemistry and Molecular Biology

UNIVERSITY: University of Texas M.D. Anderson Cancer Center

BOX: Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA

TEL: Tel: 713 792 3646

FAX: Fax: 713 790 0329.

FEATURES location/Qualifiers

1..600

/organism="Mus musculus"

/mol_type="mRNA"

/db_xref="taxon:10090"

/tissue_type="neural retina"

/dev_stage="embryonic day 14.5 post-fertilization"

/clone_lib="Mouse E14.5 retina lambda ZAP II Library"

BASE COUNT 99 a 194 c 222 g 82 t 3 others

ORIGIN

Query Match 8.2%; Score 120.4; DB 13; Length 600;

Best Local Similarity 75.1%; Pred. No. 7.5e-19;

Matches 163; Conservative 0; Mismatches 52; Indels 2; Gaps 1;

QY 705 AGCCGCGCCAGAAAGCCCAAGCAGCGGAGCGCAACCGATGACCAACTTAACCTCCGC 764

DB 336 ACCCGAGGCTCAAGGCCCAAGCGGAGCGCAACCGATGACCAACTTAACCGCGCG 395

QY 765 CTGATGCGCTGCGGCGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTAAGATC 824

DB 396 CTGAGCGGCTGCGGAGGTGTCTCTGCCCACTTCCCGATGACCGCAACTTAAGATC 455

QY 825 GAGACCT--GCGCTTCCGCCCAACTTATTTGGGCACTGACTCAGACCGTGGCATACG 882

DB 456 GAGACGCTTGGCGCTTCCGCCCAACTTATTTGGGCACTGACTCAGACCGTGGCATACG 515

QY 883 CGAGACCAAGCTTCAAGCGCCCGAGCCCGTGGCC 919

DB 516 CGAGACCAAGCTTCAAGCGCCCGAGCCCGTGGCC 552

RESULT 12
CA979119
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
TITLE
AUTHORS
JOURNAL
COMMENT

CA979119 932 bp mRNA linear EST 06-JAN-2003
AGNCOURT_11295215 NIH-MGC_164 Mus musculus cDNA clone
IMAGE:30146192 5', mRNA sequence.
CA979119
CA979119.1 GI:27511773
EST.
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 932)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabs-rt@mail.nih.gov
Tissue Procurement: Dr. David Rowe and Dr. Mina
GDN Library Preparation: Invitrogen Corp
GDN Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: NDAM0061 row: h column: 09
High quality sequence start: 16
High quality sequence stop: 640.

FEATURES		Location/Qualifiers
1.	.932	
source		/organism="Mus musculus"
		/mol_type="mRNA"
		/db_xref="taxon:10090"
		/clone="IMAGE:30146192"
		/lab_host="DH10B (phage-resistant)"
		/clone_lib="NIH MGC 164"
		/note="Vector: pCMV-SPORT6.1.cdb, Site_1: EcoRV, Site_2: NotI; Non-normalized, full-length enriched library from pooled mouse embryonic limb, maxilla and mandible, day 10.5 and 11.5 (size selected for the 0.5-1 kb fragments) Cloned directionally, priming method: Oligo-dT. cDNA enrichment: >1x bp, Average insert size 1.8k bp. Priming sequence: 5'GACTAGTCTCATATCCGAGCGGCCCTT) 3'. Tissue contributed by, David Rowe. Library constructed by ResGen, Invitrogen Corp."
BASE COUNT	153 a 349 c 276 g 153 t	1 others
ORIGIN		
Query Match	8.1%	Score 118.6; DB 14; Length 932;
Best Local Similarity	66.1%;	Pred. No. 2.5e-18;
Matches 187;	Conservative 0;	Mismatches 94; Indels 2; Gaps 1;
QY	635	GAAGCTCCGTCGCGCGCGCGGAGGCGGCAACAGCGCCCAAGACGATTGGCACTGAGCA 694
DB	464	GCAGGAACGGGGAGGCGGCGGCGAGCTCGCGCTGGGTGGGTCGAGGCTCTGTGTCATCTC 523
QY	695	GCAGCGACGAAGCCCGGCGCAAGAAAGGCCCAAGCAGCCGGAGGCCAACCAGATGCACACCT 754
DB	524	CCTGGGAGGAGTCTCTCCGCTCAAAAGCCAAAGATGCAGCGCAACCCGATGCACACCT 583
QY	755	TAACTCCGCGCTGAGTGGCTGCGCGGAGTCTCTGCGCCACTTCCCGAGTGAAGCCAAACT 814
DB	584	CAGCGTCTGCGCTGAGCGCGCTTGGCAGCGATGCTGCTGCTGCTTCCCGAGACCAACACT 643
QY	815	TACAAAGATGAGACCTCTGCGCTTGCGCCACAA--CTACATTGGGCACTGACTCAGACG 872
DB	644	CACCAAGATTGAGAGCTGCGCTTGGCTTACAAACTATCATCTGGGGCCCTGTGTCAGACA 703
QY	873	CTGGGCATAGGGGACCAACAGCTTCTAAGGCCCCCGAGCCCCCTTG 915
DB	704	CTGGCGCTGGCAGATCAAGGGCTCTCCCGGGGGCACTGGCCCCG 746

	1037 bp	DNA linear	GSS 01-SEP-2000
CNS03UJ9/c	CNS03UJ9	Tetradon nigriviridis genome survey sequence T7 end of clone 06ZM23 of library G from Tetradon nigriviridis genomic survey sequence.	
ACCESSION	AL262494.1	GI:7984120	
VERSION			
KEYWORDS	GSS; genome survey; sequence.		
SOURCE	Tetradon nigriviridis		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Acanthomorpha; Acanthopterygii; Percormorpha; Tetraodontiformes; Tetraodoncidae; Tetraodontidae; Tetraodon.		
REFERENCE	1	Roest Crolius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Fitzames,C., Wincker,P., Brottier,P., Quetier,F., Saurin,W. and Weissbach,J. Estimate of human gene number provided by genome-wide analysis using Tetradon nigriviridis DNA sequence <i>Nat. Genet.</i> 25 (2), 235-238 (2000)	
AUTHORS	JOURNAL MEDLINE PUBMED	20296633 10835645	
REFERENCE	2	Roest Crolius,H., Jaillon,O., Dasilva,C., Ozouf-Costaz,C., Fitzames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissbach,J. Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradon nigriviridis <i>Genome Res.</i> 10 (7), 939-949 (2000)	
AUTHORS	JOURNAL MEDLINE PUBMED	20359837 10899143	
REFERENCE	3	(bases 1 to 1037) Genoscope.	
AUTHORS	Direct Submission		
TITLE	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr)		
JOURNAL	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetradon nigriviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon .		
COMMENT			

FEATURES	source	Location/Qualifiers
		1..1037
		/organism="Tetrarodon nigroviridis"
		/mol_type="genomic DNA"
		/db_xref="taxon:99883"
		/clone_1="062M23"
		/clone_1b="G"
		/note="Genoscope sequence ID : COBG062AG12P1-end : T7"
BASE COUNT	205 a	254 c 267 g 308 t 3 others
ORIGIN		
Query Match	7.9%	Score 115.6; DB 29; Length 1037;
Best Local Similarity	74.7%	Pred. No. 1,4e-11;
Matches 145; Conservative	0;	Mismatches 49; Indels 0; Gaps 0;
Dy	696	CAGCGAGAAAGCCGGCCCAAGAAAGCCCAACGACCGGAGCGCAACCCGATSCAACACTT 755
Db	417	CAGCGGGGCGCAGCGCAGATGAAAGCGAACGACAGGAGCGCCACCGATGCAAACTG 358
Dy	756	AACCTCGCGCTGATGTCGCTGCGCGGTGTCCTGCGCACTTCCGGATGACGCCAAACTT 815
Db	357	AACCTCCCGCTGAGCGCGCTGAGGGGCGATCTCGCGGTGCTACCCGAGGACACAAAGCTG 298
Dy	816	ACAAAGATTCAGAGACCCGCGCTTCGCCCAACAATCATTTGGGAGCATCGATCAACGCTG 875
Db	297	ACCAAAATGTGAATCTGCGCTTTGGCCCAACTACATCTGGGGCTTTGACCGAGACTGTG 238
Dy	876	CGCATAGCGGACCA 889

[illegible]

Qy	764	GCCTGAATGCGCGTGGCGCGGTGCTGCTGCGCCACCTTCGCCGATGAGCGCCAAACTTACCAAGAT	823
Db	64	GCTCGATTTCTCTGAGGAGGAGTTCTTACCGCTATTACCGGAAGACCCAACTCACCAAGAT	123
Qy	824	CGAGACCCCTGCGCTTCGCCCAACTATCATTTTGGGCACTGACTGACGCTGCGCATAGC	883
Db	124	AGAGACCTTGCGGCTTTGCCCAACTATCATCTGGGCTCTTAGCGAAACTTTGGCGCGTGC	183
Qy	884	GGACCA 889	
Db	184	CGACCA 189	
RESULT 15			
CD282259			
LOCUS	CD282259		
DEFINITION	639271.16 NCI_CGAP_Zemb2 Danio rerio cDNA clone IMAGE:652149 5',	603 bp	mRNA linear EST 23-MAY-2003
ACCESSION	CD282259		
VERSION	CD282259.1	GI:11060035	
KEYWORDS	EST.		
ORGANISM	Danio rerio (zebrafish)		
REFERENCE	Danio rerio		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.		
TITLE	1 (bases 1 to 603)		
JOURNAL	Amundsen, C., Cachuela, N., Chen, F., Cheung, L.M., Chong, A., Murray, L., Oliva, D., Park, C., Reyes, J., Yungen, J. and Swimmer, C.		
COMMENT	Expressed sequence tags from NCI_CGAP_Zemb2, a Danio rerio embryonic library		
	Unpublished		
	Contact: Chen F.		
	Exelixis, Inc.		
	170 Harbor Way, PO Box 511, South San Francisco, CA 94083-0511, USA		
	Tel: 650 837 7000		
	Fax: 650 837 8300		
	Email: fchen@exelixis.com		
	DNA Sequencing by: Exelixis, Inc. Clone distribution information		
	can be found through the I.M.A.G.E. Consortium/LANL at:		
	http://image.llnl.gov		
	Plate: 14108 row: O column: 3		
	High quality sequence stop: 603.		
FEATURES			
Source	Location/Qualifiers		
	1..603		
	/organism="Danio rerio"		
	/mol_type="mRNA"		
	/db_xref="taxon:7955"		
	/clone="IMAGE:6521499"		
	/tissue_type="embryo"		
	/lab_host="DH10B (T1-resistant)"		
	/clone_id="NCI_CGAP_Zemb2"		
	/note="Vector: pCMV-SPORT6, ccdb; Site 1: EcoRV; Site 2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 2 kb. Constructed by J. Wang (Research Genetics, Invitrogen Corp) from tissue donated by L. Zon (Harvard University). Note: this is a NCI_CGAP Library."		
BASE COUNT	159 a 185 c 155 g 104 t		
ORIGIN			
Query Match	7.5%; Score 110.2; DB 14; Length 603;		
Best Local Similarity	74.3%; Pred. No. 2.4e-16;		
Matches 139; Conservative	0; Mismatches 48; Indels 0; Gaps 0;		
Db	705 AGCGGGCGCAAGAAAGCCCAACGACCGGGAGGCGCAACGCATGACCAACTTAATCGCGC	764	
325 AACCGCAGGCTGAAGAGCCCAACGACCGCGAGAGGAACAGATGCACAACTTAAAGAGCA	384		
Qy	765 CTGATGCGCTGCGCGGTGCTGCTGCGCCACTTTCGCCGATGAGCGCCAACTTACCAAGATC	824	
Db	385 TTGATGCTTTGAAGAGGCTCTGCTGCGCTTCTCTGACGACAAAGCTGACCAAAATT	444	
Qy	825 GAGACCTGCGCTTGCCCAACTACTATTTGGGCACTGACTGACGCTGCGCATAGC	884	

Db	445	GAGACTCTCGCTTGCCTCACAACCTACATCTGCGGCACTTTCGAGACCATCCGGAATCGCA	504
Qy	885	GACCAACA	891
Db	505	GACCAAGA	511

Search completed: January 26, 2004, 22:05:05
Job time : 3312 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 18:54:03 ; Search time 5306 Seconds

(without alignments)
11256.701 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460

Sequence: 1 gcagtgacgagagagagcag.....agagtgacccaatccagtggt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : GenEmbl.*

```
1: gb_da:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_da:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1409	96.5	1491	6 A91167	A91167 Sequence 1
2	1409	96.5	1491	6 BD023626	BD023626 Polypepti
3	1409	96.5	1491	10 RNRELAXT	Y10619 R.norvegicu
4	1060	72.6	258815	2 AC127817	AC127817 Rattus no
5	92	6.3	861	6 AX698801	AX698801 Sequence
6	92	6.3	861	10 MMU76208	U76208 Mus musculu
7	92	6.3	1861	10 AF364300	AF364300 Mus muscu
8	92	6.3	5567	10 MMAT4B	Y09167 M.musculus
9	92	6.3	215050	2 AC109783	AC109783 Mus muscu
10	92	6.3	215050	2 AC127417	AC127417 Mus muscu
11	35	2.4	2370	9 BC036847	BC036847 Homo sapi
12	35	2.4	6123	9 AF303002	AF303002 Homo sapi
13	35	2.4	91531	2 AC079846_3	Continuation (4 of
14	35	2.4	179697	9 AC023886	AC023886 Homo sapi
15	32	2.2	1330	9 HSA133776	A113776 Homo sapi
16	32	2.2	5340	9 AF234829	AF234829 Homo sapi
17	32	2.2	165110	9 AL450311	AL450311 Human DNA
18	32	2.2	173341	2 AC021954	AC021954 Homo sapi
19	26	1.8	790	6 AX548297	AX548297 Sequence
20	26	1.8	170896	2 AC011010	AC011010 Homo sapi
21	25	1.7	25	6 A91170	A91170 Sequence 4
22	25	1.7	25	6 BD023629	BD023629 Polypepti
23	23	1.6	735	10 MMU67776	U67776 Mus musculu
24	23	1.6	738	6 AR308548	AR308548 Sequence
25	23	1.6	770	5 AF123884	AF123884 Gallus ga
26	23	1.6	790	5 GGA012659	AJ012659 Gallus ga
27	23	1.6	932	10 MMU63841	Y09166 M.musculus
28	23	1.6	1315	6 MMU63841	U63841 Mus musculu
29	23	1.6	1333	6 AR023715	AR023715 Sequence
30	23	1.6	1333	6 AR225848	AR225848 Sequence
31	23	1.6	1341	5 AF109014	AF109014 Gallus ga
32	23	1.6	1385	10 MMU76207	U76207 Mus musculu
33	23	1.6	1412	10 MMAT4B	Y07621 M.musculus
34	23	1.6	1880	5 AF303000	AF303000 Gallus ga
35	23	1.6	10393	10 AF303001	AF303001 Mus muscu
36	23	1.6	71538	2 AC118243	AC118243 Mus muscu
37	23	1.6	123855	2 AC102600	AC102600 Mus muscu
38	23	1.6	149268	2 AC124395	AC124395 Mus muscu
39	23	1.6	262798	2 AC111702	AC111702 Rattus no
40	22	1.5	215591	2 AC102159	AC102159 Mus muscu
41	21	1.4	21	6 AX553597	AX553597 Sequence
42	21	1.4	21	6 AX553645	AX553645 Sequence
43	21	1.4	310	6 BC052235	BC052235 Homo sapi
44	21	1.4	1058	9 BC052235	BC052235 Homo sapi
45	21	1.4	1164	5 GENEURODL	Y09596 G.gallus mr

ALIGNMENTS

```
RESULT 1
LOCUS      A91167
DEFINITION Sequence 1 from Patent WO9827206.
ACCESSION  A91167
VERSION    A91167.1 GI:6740202
KEYWORDS
SOURCE
ORGANISM   Rattus sp.
            Rattus sp.
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 1491)
AUTHOR    Icard-Liepkalns, C., Maliet, J. and Corresponding, N.A.
JOURNAL   Patent: WO 9827206-A 1 25-JUN-1998;
```


	PR	19-DEC-1996 FR	96/15651	
	PI	CHRISTINE ICARD LIEBKUNLS JACQUES MALET, PHILIPPE RAVASSARD PC		
	C07K4/47,A6IK31/711,A6IK35/76,A6IK38/00,A6IK48/00,A6I25/00,PC			
	CI2N15/09,			
	PC	CI2N15/00,A6IK37/02		
	CC	Strandedness: Single;		
	CC	Topology: Linear;		
	FH	Key	Location/Qualifiers.	
FEATURES				
Source			1..1491	
		/organism="Rattus sp."		
		/mol type="genomic DNA"		
		/db xref="taxon:10118"		
BASE COUNT				
ORIGIN				
	307 a	_487 c	413 g	284 t
Query Match		96.5%; Score 1409; DB 6; Length 1491;		
Best Local Similarity		99.9%; Pred. No. 0;		
Matches 1459; Conservative	0; Mismatches	1; Indels	0; Gaps	0
OY	1	GCAGGTACCGAGAAGGACAGTCCCTGGGCCCCCCTTGCTGATTGGCCCGTCGACAAGCA	60	
Db	1	GCAGGTACCGAGAAGGACAGTCCCTGGGCCCCCCTTGCTGATTGGCCCGTCGACAAGCA	60	
OY	61	Gcagcccgagacggcacgcgtccctggtccccggagacagataaaagggtgcacaaggagacaca	120	
Db	61	Gcagcccgagacggcacgcgtccctggtccccggagacagataaaagggtgcacaaggagacaca	120	
OY	121	CGATTAGCAGCTCAGAAAGTCCCTCGGTGCTCATCCACTGCAAGAGCGCGAAGACCCT	180	
Db	121	CGATTAGCAGCTCAGAAAGTCCCTCGGTGCTCATCCACTGCAAGAGCGCGAAGACCCT	180	
OY	181	CCGAGCTTCTTTGCTGCTCCTCCAGACGCAATTTACTTCACGCGAGGCGCTGACGCTCAG	240	
Db	181	CCGAGCTTCTTTGCTGCTCCTCCAGACGCAATTTACTTCACGCGAGGCGCTGACGCTCAG	240	
OY	241	CAAAACCTTCGAAGCCAGACGAGAGGGGTTCAGCTATCCAACGCTGCTTAATCTTGACACCC	300	
Db	241	CAAAACCTTCGAAGCCAGACGAGAGGGGTTCAGCTATCCAACGCTGCTTAATCTTGACACCC	300	
OY	301	GCAGCTCTGTCTTTTGGAGCCCGAGATTAATAGAGTAACTTTAGAACTCTCAAAGGG	360	
Db	301	GCAGCTCTGTCTTTTGGAGCCCGAGATTAATAGAGTAACTTTAGAACTCTCAAAGGG	360	
OY	361	TAGAAAGAGGGAGTGGGTGGGCGTACTCTAGTCCGCGTGAAGTGAACCTTAAGTCAGAG	420	
Db	361	TAGAAAGAGGGAGTGGGTGGGCGTACTCTAGTCCGCGTGAAGTGAACCTTAAGTCAGAG	420	
OY	421	ACTGTCAACACCCCCCTTTCATTTTTTCCCACTCAAGATGGCGCTCATCCCTTGATG	480	
Db	421	ACTGTCAACACCCCCCTTTCATTTTTTCCCACTCAAGATGGCGCTCATCCCTTGATG	480	
OY	481	CGCCCAACATCCAAGTGTCCCAAGAGAACCCAGCAACCTTTCCCGAGGCTCGAGCACAG	540	
Db	481	CGCCCAACATCCAAGTGTCCCAAGAGAACCCAGCAACCTTTCCCGAGGCTCGAGCACAG	540	
OY	541	AAGTGTCAAGTTCCAATTCACCCCAACTTAGCCCACTTCGTACCGAGGGACTGCTCG	600	
Db	541	AAGTGTCAAGTTCCAATTCACCCCAACTTAGCCCACTTCGTACCGAGGGACTGCTCG	600	
OY	601	AAGCAAGAAAGCAGGTACTGTCGCGAGGGAATCAAGGAAGCTCCGTGCGCGCCGAGAGGC	660	
Db	601	AAGCAAGAAAGCAGGTACTGTCGCGAGGGAATCAAGGAAGCTCCGTGCGCGCCGAGAGGC	660	
OY	661	GCAACAGGCCCAAGACGAGTGGGACTGAGGAAGCAGGACGAGACCGGCGGCAAGAAAG	720	
Db	661	GCAACAGGCCCAAGACGAGTGGGACTGAGGAAGCAGGACGAGACCGGCGGCAAGAAAG	720	
OY	721	CCAAAGACCGGAGGCGCAACCGCATGACCAACTTTAACTCCGCGTGAATGCGTGCAG	780	
Db	721	CCAAAGACCGGAGGCGCAACCGCATGACCAACTTTAACTCCGCGTGAATGCGTGCAG	780	
OY	781	GTGTCTGCTCCACCTTCCCGAGTGAAGCCCAAACTTAAGATCGAGACCTTGCGCTTCG	840	

Db	781	GTGTCTGTGCCACCTTCCCGATGACCGCAAACTTACAAABATGAGAGCCTTGGCTTCG	840
QY	841	CCCAACAATCAATTTGGGCACTGACTGACGCTGCGCATAGCGGACCAAGCTTTACG	900
Db	841	CCCAACAATCAATTTGGGCACTGACTGACGCTGCGCATAGCGGACCAAGCTTTACG	900
QY	901	GCCCCGAGCCCCCTGTGTGCCCTGTGTGGGAGAGCTGGGAAAGCCCGGAGGGGGCTTCAGCGGCG	960
Db	901	GCCCCGAGCCCCCTGTGTGCCCTGTGTGGGAGAGCTGGGAAAGCCCGGAGGGGGCTTCAGCGGCG	960
QY	961	ACTGGGGCTCATCTACTCCCCAGTTTTCCCAAGCTGTGATGAGCTGAGCCCCCAGAGCTCAT	1020
Db	961	ACTGGGGCTCATCTACTCCCCAGTTTTCCCAAGCTGTGATGAGCTGAGCCCCCAGAGCTCAT	1020
QY	1021	TGAGAGAGTTCCCTGGGCTTGAGAGGTGCCAGCTCCCAATCTGTGTGCTCCCGGACACC	1080
Db	1021	TGAGAGAGTTCCCTGGGCTTGAGAGGTGCCAGCTCCCAATCTGTGTGCTCCCGGACACC	1080
QY	1081	TGTGTGTTCTAGACTTTCTTGTGAAGGGGCCAAACAGAGCCCTGGGCGGTGGGCGCTGGGAG	1140
Db	1081	TGTGTGTTCTAGACTTTCTTGTGAAGGGGCCAAACAGAGCCCTGGGCGGTGGGCGCTGGGAG	1140
QY	1141	AAAGGAGAGGAGTGAAGAGCTGTCTGAATATGGAAGTATGAGGCACTCGAGACTCTGCG	1200
Db	1141	AAAGGAGAGGAGTGAAGAGCTGTCTGAATATGGAAGTATGAGGCACTCGAGACTCTGCG	1200
QY	1201	CCCTTCTGGCTTTTCATTAGTCAGAGTCCCTGATTTTAAACAGAGATTGCAAGTTCCTTCT	1260
Db	1201	CCCTTCTGGCTTTTCATTAGTCAGAGTCCCTGATTTTAAACAGAGATTGCAAGTTCCTTCT	1260
QY	1261	GCTGTGGGTGACAAAGGATTTGCAAGCTGATCTCTTTAACCCCTCTCAGTGTGCC	1320
Db	1261	GCTGTGGGTGACAAAGGATTTGCAAGCTGATCTCTTTAACCCCTCTCAGTGTGCC	1320
QY	1321	ACCTCAAACTCCGCTCCAGCAGAGGAGACCCGTAAGCTAAATATGTTGGGAGACTCC	1380
Db	1321	ACCTCAAACTCCGCTCCAGCAGAGGAGACCCGTAAGCTAAATATGTTGGGAGACTCC	1380
QY	1381	ATACTTCCTGGTGACTCCGCCCTTTTCAATATGCGGGCCCTCCAAACACGCTTCTCC	1440
Db	1381	ATACTTCCTGGTGACTCCGCCCTTTTCAATATGCGGGCCCTCCAAACACGCTTCTCC	1440
QY	1441	AGAGTGAACCTAATCAAGTGT 1460	
Db	1441	AGAGTGAACCTAATCAAGTGT 1460	
RESULT 3			
RNRELAXT			
LOCUS		1491 bp	linear
DEFINITION		R.norvegicus mRNA for transcriptional regulator, Relax.	ROD 06-MAY-1997
ACCESSION		V10619	
VERSION		V10619.1	GI:2072737
KEYWORDS		Relax; transcriptional regulator.	
SOURCE		Rattus norvegicus (Norway rat)	
ORGANISM		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.	
REFERENCE			
AUTHORS		1	
TITLE		Ravassard, P., Chatail, F., Mallet, J. and Icard-Liepkalns, C.	
JOURNAL		Relax, a novel rat bHLH transcriptional regulator transiently	
MEDLINE		expressed in the ventricular proliferating zone of the developing	
PUBMED		central nervous system	
REFERENCE		J. Neurosci. Res. 48 (2), 146-158 (1997)	
AUTHORS		2 (bases 1 to 1491)	
TITLE		Ravassard, P.	
JOURNAL		Direct Submission	
MEDLINE		Submitted (20-JUN-1997) P. Ravassard, CNRS UMR 9923, Bat. CERV1,	
PUBMED		Hopital de la Pltie Salpatriere, 83 Bd. de l'Hopital, F-75013	

FEATURES Paris, FRANCE
Location/Qualifiers
1.1491
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Mistar"
/db_xref="taxon:10116"
/clone_1lb="RT-PCR"
459..1103
/function="transcriptional regulator"
/codon_start=1
/product="Relax"
/protein_id="CAA71630.1"
/db_xref="GI:2072738"
/db_xref="SPTREMBL:O08718"
/translation="MAPPPLDAPITOVSETOQPPFGASDHEVLSNSTPSPITVPR
DLSEAGDCKRSTRLKARRGGRNPKSELISKORRRKKANDRENNHNLISA
IDLRGLVLPFPDPAKLTKEITLRFANHYIMALTQRLRLADHSFYGPPEPVPVCGELGS
PGSGSGDMGSIYSPVQAGSISPTASLEFPGLQVPSPSCLIPQLTVFSDFL"

BASE COUNT 307 a 487 c 413 g 284 t

ORIGIN

Query Match 96.5%; Score 1409; DB 10; Length 1491;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1459; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GCAGGTAGGAGAGGAGGAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 60
1 GCAGGTAGGAGAGGAGGAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 60
61 GCAGCCCGGAGGAGGAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 120
61 GCAGCCCGGAGGAGGAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 120
61 GCAGCCCGGAGGAGGAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 120
121 CGATTAGCAGCTCAGAAAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 180
121 CGATTAGCAGCTCAGAAAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 180
121 CGATTAGCAGCTCAGAAAGTCCCTGGGCCCCGTTGATTTGGCCCGTGCACAGGCA 180
181 CGAGGTTCTTTGCTGCTCCCTCCAGACGCAATTACTCCAGGCGGCGCTGCAGTCA 240
181 CGAGGTTCTTTGCTGCTCCCTCCAGACGCAATTACTCCAGGCGGCGCTGCAGTCA 240
181 CGAGGTTCTTTGCTGCTCCCTCCAGACGCAATTACTCCAGGCGGCGCTGCAGTCA 240
241 CAAAATTGCAAGCGAGAGGAGGTTCACTACCGCTCTTGAACCTGACACACCC 300
241 CAAAATTGCAAGCGAGAGGAGGTTCACTACCGCTCTTGAACCTGACACACCC 300
301 GCAGCTCTCTGTTCTTTTGAAGCCCGAGTAAGTAACTTTAGGAACCTTCAAGG 360
301 GCAGCTCTCTGTTCTTTTGAAGCCCGAGTAAGTAACTTTAGGAACCTTCAAGG 360
361 TAGAAGAGGAGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 420
361 TAGAAGAGGAGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 420
421 ACTGTCAACACCCCTTCCATTTTTCACACCTCAGAGTGGGCTCATCCCTGGAG 480
421 ACTGTCAACACCCCTTCCATTTTTCACACCTCAGAGTGGGCTCATCCCTGGAG 480
481 CGGCCACATCCAAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTTCGAG 540
481 CGGCCACATCCAAAGTGTCCCAAGAGACCCAGAACCTTTCCCGAGGCTTCGAG 540
541 AATGTCTCAATTTCACATCCCACTGAGCCCACTTCGTACCGAGGAGTGTCTCC 600
541 AATGTCTCAATTTCACATCCCACTGAGCCCACTTCGTACCGAGGAGTGTCTCC 600
601 AACGAGAGGAGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 660
601 AACGAGAGGAGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 660
661 GCAACAGGAGGAGGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 720
661 GCAACAGGAGGAGGAGTGGGTGGGCTTACTTATGTCCTCCGCTGAGTGAACCTTGA 720

721 CCAACGACCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
721 CCAACGACCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 780
781 GTGTCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
781 GTGTCTTCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
841 CCCCAACTTACTTATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
841 CCCCAACTTACTTATTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
901 GCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
901 GCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 960
961 ACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
961 ACTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
1021 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1021 TGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1080
1081 TGGTGTCTCAGACCTTTTGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1081 TGGTGTCTCAGACCTTTTGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1140
1141 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
1141 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
1201 CCTTCTGAGCTTCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
1201 CCTTCTGAGCTTCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
1201 CCTTCTGAGCTTCAATTAATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1260
1261 GCTGTGCGTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
1261 GCTGTGCGTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
1261 GCTGTGCGTGCAAAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1320
1321 ACCTCAATCCCGCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
1321 ACCTCAATCCCGCTCCCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1380
1381 ATACTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1381 ATACTTCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1440
1441 AGAGTACCTAATCCAGTGT 1460
1441 AGAGTACCTAATCCAGTGT 1460

RESULT 4
AC127817 258815 bp DNA linear HTG 19-NOV-2002
LOCUS Rattus norvegicus clone CH230-259616, WORKING DRAFT SEQUENCE, 3
DEFINITION unordered pieces.
AC127817
AC127817.3 GI:25077905
VERSION HTG: HTGS PHASE1; HTGS DRAFT; HTGS_FULFILLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 258815)
REFERENCE 1 (bases 1 to 258815)
AUTHORS Muzny,D.,Marie., Metzger,M.,Lee., Abiraman,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsdorfs,S., Amin,A., Anguiano,D.,
Anyalebech,V., Ayodeji,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Bernhamed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,

Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A., Chavez, J., Chen, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., d'Souza, L., Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Derramo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Gantla, R., Garcia, A., Garner, T., Garza, M., Gebregioris, B., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, K., Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J., Hollins, B., Howells, S., Huylk, S., Hume, J., Idledit, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyet, A., Kaprathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowitz, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Louised, H., Lozano, R.J., Lu, X., Ma, J., Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munday, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokwemeh, O., Okumu, G., Olarnunagoon, A., Pal, S., Parke, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Pu, M., Pu, M., Quito, J., Rachlin, E., Reeves, K., Reiter, M., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rivers, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatman, S., Shen, H., Sherry, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Stead, M., Strong, R., Sutton, A., Svatek, A., Taber, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uemami, K., Valas, R., Vera, V., Villalana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Wilson, R., Wleczek, R., Woodson, H., Wolley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, J., Zhou, S., Zhao, S., Dunn, D., Von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

TITLE
Unpublished
2 (bases 1 to 258815)
Worley, K.C.
REFERENCE
Submitted (19-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 258815)
Rat Genome Sequencing Consortium.
AUTHORS
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
TITLE
Direct Submission
JOURNAL
Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
COMMENT
On Nov 19, 2002 this sequence version replaced gi:23912578.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine

Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GZXS
Center clone name: CH230-259616
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 224747 bases at least Q40
Consensus quality: 227981 bases at least Q30
Consensus quality: 229752 bases at least Q20
Estimated insert size: 228243; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbankdraft_data.html).
* NOTE: This is a "working draft" sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 255979: contig of 255979 bp in length
* 255980 256079: gap of unknown length
* 256080 257349: contig of 1270 bp in length
* 257350 257449: gap of unknown length
* 257450 258815: contig of 1366 bp in length.
Location/Qualifiers
1. 258815
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-259616"
1. 1609
/note="wgs end extension
clone end:5p6"
complement(2865..3659)
/note="clone boundary
clone_end:5p6
site:
end_sequence:BZ169379"
224098..225157
/note="clone boundary
clone_end:T7
site:
end_sequence:BZ169378"
BASE COUNT 59351 a 56888 c 58131 g 56821 t 27624 others
ORIGIN
Query Match 72.6%; Score 1060; DB 2; Length 258815;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1110; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 350 CCTCAAGGGGATGAAGAGGGGAGTGGGCGTACTGATCCCGGTGAGTACCT 409
DB 247751 CCTCAAGGGGATGAAGAGGGGAGTGGGCGTACTGATCCCGGTGAGTACCT 247810
QY 410 CTAAGTCAGAGACTGTCAACACCCCTTCATTTTTCCTCAACCTCAGATGCGCTCA 469
DB 247811 CTAAGTCAGAGACTGTCAACACCCCTTCATTTTTCCTCAACCTCAGATGCGCTCA 247870
QY 470 TCCCTTGATGCGCCCACTCAATCCAGAGACCCAGCAACCTTTCCGAGGC 529
DB 247871 TCCCTTGATGCGCCCACTCAATCCAGAGACCCAGCAACCTTTCCGAGGC 247930
QY 530 CTCGAGCAGCAAGGCTCAATCCAGAGACCCAGCAACCTTTCCGAGGC 589
DB 247931 CTCGAGCAGCAAGGCTCAATCCAGAGACCCAGCAACCTTTCCGAGGC 247990
QY 590 GGACTGCTCGAAGCAGAGAGGTGATGCGGAGGACATGAGAGACTTCGTCGCG 649


```

/db xref="SWISS-PROT:P70661"
/translation="MAPHPLDALTIQVSPETQPPGASDHEVLSNSTEPPTLIPR
DCEAEVGDGCRGTRKRLRGRGRPKSEIALSKORSRGRKRRKRNHNHNSA
LDALRGVLPPTPDAAKTETETLRAHNYMALTOTLRADPSFKGPDPVCGELGS
PFGSGNDMGSIYSFVSQAGNISPTASLEFPGLQVPSSPITLLGALVSDPL"
BASE COUNT      1271 a 1549 c 1564 g 1183 t
ORIGIN
Query Match      6.3%; Score 92; DB 10; Length 5567;
Best Local Similarity 100.0%; Pred. No. 1,1e-42;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGATGAGCCCAACTTACAAG 821
    |||||
Db 5226 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGATGAGCCCAACTTACAAG 5285

Qy 822 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 853
    |||||
Db 5286 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 5317

RESULT 9
AC109783/c      138070 bp      DNA      linear      HTG 07-FEB-2002
LOCUS          Mus musculus clone RP23-121F10, WORKING DRAFT SEQUENCE, 17
DEFINITION     unordered pieces.
ACCESSION      AC109783
VERSION        AC109783.1 GI:18581594
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      1 (bases 1 to 138070)
AUTHORS        McCombie,W.R., Baker,J.P., Balija,V., Dedhia,N.N., de la
               Bastide,M., Katzenberger,F., Kuit,K., King,L., Kirchoff,K.A.,
               Miller,B., Muller,S., Nascimento,L.U., O'Shaughnessy,A.L.,
               Preston,R.R., Santos,L., Spiegel,L.A., Palmer,L., Yang,C. and
               Zuberi,T.
TITLE          Mouse Genomic Sequence
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 138070)
AUTHORS        McCombie,W.R.
TITLE          Direct Submissiion
JOURNAL        Submitted (07-FEB-2002) Lita Annenberg Hazen Genome Sequencing
               Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
               Harbor, NY 11724, USA
COMMENT        -----
               Center: Lita Annenberg Hazen Genome Center, Cold Spring Harbor
               Laboratory
               Center code: CSHL
               Web site: http://www.cshl.org/geneseq
               Contact: mcombie@cshl.org
               Project Information
               Project name: RP23-121F10
               Clone name: RP23-121F10
               Insert size: 173000; agarose-fp
               Insert size: 141616; sum-of-ctnigs
               Quality coverage: 4.00 in Q20 bases; agarose-fp
               Quality coverage: 3.70 in Q20 bases; sum-of-ctnigs
               * NOTE: This is a 'working draft' sequence. It currently
               * consists of 17 contigs. The true order of the pieces
               * is not known and their order in this sequence record is
               * arbitrary. Gaps between the contigs are represented as
               * runs of N, but the exact sizes of the gaps are unknown.
               * This record will be updated with the finished sequence
               * as soon as it is available and the accession number will
               * be preserved.
               *
               * 1 17785: contig of 17785 bp in length
               * 17786 17874: gap of unknown length
               * 17875 30356: contig of 12482 bp in length
               * 30357 30444: gap of unknown length

/db xref="taxon:10090"
/mol type="genomic DNA"
/db xref="taxon:10090"
/clone="RP23-121F10"

FEATURES
source
1..138070
location/Qualifiers
1..138070
/organism="Mus musculus"
/mol type="genomic DNA"
/db xref="taxon:10090"
/clone="RP23-121F10"

BASE COUNT      35731 a 33657 c 33303 g 33954 t 1425 others
ORIGIN
Query Match      6.3%; Score 92; DB 2; Length 138070;
Best Local Similarity 100.0%; Pred. No. 9.7e-43;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 762 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGATGAGCCCAACTTACAAG 821
    |||||
Db 110917 GCGCTGATGCGCTGCGCGGTCTCTGCCACCTTCCGATGAGCCCAACTTACAAG 110858

Qy 822 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 853
    |||||
Db 110857 ATCGAGACCTTGCGCTTCCGCCCAACTTACAT 110826

RESULT 10
AC127417      215050 bp      DNA      linear      HTG 19-OCT-2002
LOCUS          Mus musculus chromosome UNK clone RP23-459M2, WORKING DRAFT
DEFINITION     SEQUENCE, 7 unordered pieces.
ACCESSION      AC127417
VERSION        AC127417.2 GI:24137619
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FILLTOP.
SOURCE         Mus musculus (house mouse)
ORGANISM       Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE      1 (bases 1 to 215050)
AUTHORS        Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
               McPherson,J.D. and Waterston,R.H.
TITLE          The sequence of Mus musculus clone
JOURNAL        Unpublished
REFERENCE      2 (bases 1 to 215050)
AUTHORS        McPherson,J.D. and Waterston,R.H.
TITLE          Direct Submissiion
JOURNAL        Submitted (15-JUN-2002) Genome Sequencing Center, 4444 Forest Park
               Parkway, St. Louis, MO 63108, USA
REFERENCE      3 (bases 1 to 215050)

```

1120

Email: cgabps-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc_mgc@hgrl.nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S. M., Benjamin, B.,
Blakesley, R. W., Bouffard, G. G., Breen, K., Binkley, C., Brooks, S.,
Dierich, N. L., Granite, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S. L., Karlins, E., Kwong, P., Latic, P., Legaspi, R. C.,
Maduro, Q. L., Masello, C., Marker, B., Mastrian, S. D., McCloskey, J. C.,
McDowell, J., Pearson, R., Stancitrop, S., Thomas, P. J., Touchman, J. W.

Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/URL at: <http://image.llnl.gov>
Series: IRAC Plate: 78 Row: K Column: 16.
Location/Qualifiers

FEATURES

source

1. 2370
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_id="MGC:46562 IMAGE:5247719"
/cissue_type="Brain, fetal, whole pooled"
/clone_lib="NIH MGC_121"
/lab_host="DH10B"
/note="Vector: PCMV-SPORE"

gene

CDS

1. 2370
/gene="NEUROG2"
/note="synonyms: Atoh4, Math4a, ngn-2, MGC46562"
/db_xref="GeneID:63973"
/db_xref="MIM:606624"
328..1146
/codon_start=1
/product="neurogenin 2"
/protein_id="AAH36847.1"
/db_xref="GI:22477417"
/db_xref="LocusID:63973"
/translation="MFEVKSETLKEKEEDVVLGSGAPALALPTPLSSADEEBEE
PASGGARRQGAEGAGQARGVAGAECCRPRLGLVHDCRRSRARAAVRSKAT
AETVQRICKTRRLKANNRRNMNINADLREVLPTPEPAKTKLETLPFANN
IWALETTLADHCGGGLPGALFSEAVLISPGASALSSSGSPSPASTSCSTN
SPAPSSSVSNSTSPYSCITLSPSPASPMQPPPKHRYAPHLPIARDCI"

BASE COUNT

580 a 654 c 616 g 520 t

ORIGIN

Query Match

Best Local Similarity 100.0%; Score 35; DB 9; Length 2370;

Matches

35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853

DB

775 AAGATCGAGACCTCGGCTTGGCCCACTACAT 809

RESULT 12

AF303002

LOCUS AF303002 6123 bp DNA linear PRI 13-NOV-2001
DEFINITION Homo sapiens neurogenin 2 gene, partial cds.
ACCESSION AF303002
VERSION AF303002.1 GI:11875763

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 6123)
Simmons, A.D., Horton, S., Abney, A.L. and Johnson, J.E.
Neurogenin2 expression in ventral and dorsal spinal neural tube
progenitor cells is regulated by distinct enhancers
Dev. Biol. 229 (2), 327-339 (2001)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 6123)
Simmons, A.D., Horton, S., Abney, A.L. and Johnson, J.E.
Direct Submision
Submitted (06-SEP-2000) Center for Basic Neuroscience - N44.146, UT
Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, TX
75390-9111, USA

TITLE

Location/Qualifiers
1. 6123
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES

source

Location/Qualifiers
1. 6123
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

mRNA
CDS
1. 754
/product="neurogenin 2"
/note="NM2"
/codon_start=2
/product="neurogenin 2"
/protein_id="AA040770.1"
/db_xref="GI:11875764"
/translation="ASPALALPTPLSSADEEBEEPASGGARRQGAEGARG
VAAGCGRRRLGLVHDCRRSRARAAVRSKATVQRICKTRRLKANNRRN
MNNALADLREVLPTPEPAKTKLETLPFANNIWALETTLADHCGGGLP
GALFSEAVLISPGASALSSSGSPSPASTSCSTNPPSPSSSVSNSTSPYSCITLSP
ASPMQPPPKHRYAPHLPIARDCI"

BASE COUNT

1484 a 1536 c 1507 g 1596 t

ORIGIN

Query Match 2.4%; Score 35; DB 9; Length 6123;
Best Local Similarity 100.0%; Pred. No. 2e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853

DB

383 AAGATCGAGACCTCGGCTTGGCCCACTACAT 417

RESULT 13

AC079846_3/c

WPCOMMENT

Sequence split into 4 fragments LOCUS AC079846 Accession AC079846

Fragment Name

Begin End

AC079846_1

1 110000

AC079846_2

100001 210000

AC079846_3

200001 310000

AC079846_4

300001 391531

Continuation (4 of 4)

of AC079846 from base 300001 (AC079846 Homo sapiens chromosome 3 c

Query Match

2.4%; Score 35; DB 2; Length 91531;

Best Local Similarity

100.0%; Pred. No. 1.8e-08;

Matches

35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

819 AAGATCGAGACCTCGGCTTGGCCCACTACAT 853

DB

86995 AAGATCGAGACCTCGGCTTGGCCCACTACAT 86961

RESULT 14

AC023886/c

LOCUS AC023886 179697 bp DNA linear PRI 20-MAR-2002
DEFINITION Homo sapiens BAC clone RP11-402J6 from 4, complete sequence.
ACCESSION AC023886
VERSION AC023886.7 GI:19482381

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 179697)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

2 (bases 1 to 179697)
Radionenko, M. and Abbott, A.
The sequence of Homo sapiens BAC clone RP11-402J6
Unpublished (2001)

TITLE

3 (bases 1 to 179697)
Waterston, R.H.
Direct Submision
Submitted (18-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

JOURNAL

AUTHORS

TITLE

REFERENCE

4 (bases 1 to 179697)
Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA

AUTHORS Waterston, R.H.
TITLE Direct Submission
JOURNAL Submitted (15-MAR-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
REFERENCE 5 (bases 1 to 179697)
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (20-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT On Mar 15, 2002 this sequence version replaced gi:17352441.

----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: <http://genome.wustl.edu/gsc>
Contact: sapiens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH040206

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:
The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Cataneese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>
VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is AC004049, 2000 bp overlap, the clone sequenced to the right is RP11-14886. Actual end of this clone is at base position 179697 of RP11-40206.

FEATURES

source

1. 179697
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="4"
/map="4"
/clone="RP11-40206"
/clone_1kb="RPCT-11"
1. 749
/rpt_family="L1"
repeat_region
734. 3067
/rpt_family="L1"
repeat_region
3104. 3444
/rpt_family="MALR"
repeat_region
3447. 4289
/rpt_family="L1"
repeat_region
4290. 4369
/rpt_family="L1"
repeat_region
4370. 4797
/rpt_family="MALR"
repeat_region

/rpt_family="L1"
repeat_region
5022. 5194
/rpt_family="L2"
repeat_region
5214. 5628
/rpt_family="L1"
repeat_region
5710. 5729
/rpt_family="TA)n"
repeat_region
5749. 5863
/rpt_family="TA)n"
repeat_region
5883. 6263
/rpt_family="ERV.L"
repeat_region
9901. 10151
/rpt_family="ERV.L"
repeat_region
10152. 10310
/rpt_family="L1"
repeat_region
10311. 10400
/rpt_family="Mariner"
repeat_region
10401. 10816
/rpt_family="L1"
repeat_region
10967. 12362
/rpt_family="L1"
repeat_region
12366. 12653
/rpt_family="Alu"
repeat_region
12695. 13940
/rpt_family="L1"
repeat_region
13965. 14272
/rpt_family="Alu"
repeat_region
14274. 14795
/rpt_family="ERV.L"
repeat_region
14806. 15192
/rpt_family="L1"
repeat_region
21504. 22112
/rpt_family="L1"
repeat_region
22131. 23166
/rpt_family="L1"
repeat_region
23246. 23386
/rpt_family="MALR"
repeat_region
23511. 23654
/rpt_family="MIR"
repeat_region
23655. 25486
/rpt_family="L1"
repeat_region
25487. 25636
/rpt_family="MIR"
repeat_region
25814. 26124
/rpt_family="Alu"
repeat_region
26444. 26592
/rpt_family="MIR"
repeat_region
27423. 27713
/rpt_family="Alu"
repeat_region
27716. 27739
/rpt_family="AT_rich"
repeat_region
27773. 27795
/rpt_family="AT_rich"
repeat_region
28020. 28429
/rpt_family="ERV.L"
repeat_region
29646. 29668
/rpt_family="AT_rich"
repeat_region
29727. 29871
/rpt_family="L1"
repeat_region
29875. 30114
/rpt_family="Alu"
repeat_region
30134. 30267
/rpt_family="L1"
repeat_region
30268. 30549
/rpt_family="L1"
repeat_region
30550. 30883
/rpt_family="MALR"
repeat_region
32461. 32791
/rpt_family="MALR"
repeat_region
32792. 33487
/rpt_family="L1"
repeat_region
33488. 33847
/rpt_family="MALR"
repeat_region

```

repeat_region      33848..34354
                    /rpt_family="L1"
repeat_region      34358..34379
                    /rpt_family="(CATATA)n"
repeat_region      34381..34502
                    /rpt_family="Alu"
repeat_region      34503..34556
                    /rpt_family="L1"
repeat_region      34557..34859
                    /rpt_family="Alu"
repeat_region      34860..35041
                    /rpt_family="L1"
repeat_region      35219..35676
                    /rpt_family="L2"
repeat_region      35974..36042
                    /rpt_family="L2"
repeat_region      36833..37144
                    /rpt_family="Alu"
misc_feature       36833..36848
                    /note="match to EST AY537116 (NID:g4451251) t01sh09.x1"
repeat_region      37147..37410
                    /rpt_family="Alu"
repeat_region      38154..38424
                    /rpt_family="Alu"
repeat_region      39941..40230
                    /rpt_family="Alu"
misc_feature       39941..39956
                    /note="match to EST AA75240 (NID:g2834574) ad18a05.s1"

```

```

Query Match      2.4%; Score 35; DB 9; Length 179697;
Best Local Similarity 100.0%; Pred. No. 1.7e-08;
Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      819 AAGATGAGACCTGCGCTTGGCCGACACTACAT 853
Db      55815 AAGATGAGACCTGCGCTTGGCCGACACTACAT 55781

```

```

RESULT 15
LOCUS      HSA133776              1330 bp      DNA              linear      PRI 19-JUN-1999
DEFINITION Homo sapiens gene for neurogenin 3.
ACCESSION  AJ133776
VERSION     AJ133776.1 GI:5123782
KEYWORDS    bHLH transcription factor; neurogenesis; neurogenin 3; ngn3 gene.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
            Ravassard,P., Icard-Liepkalns,C., Wiard,L., Julien,J.P. and
            Mallet,J.
            The human neurogenin 3 homolog maps to chromosome 10q21.3 and its
            expression pattern is identical to that of its murine counterpart
            Unpublished
            2 (bases 1 to 1330)
JOURNAL     Direct Submission
AUTHORS     Ravassard,P.
TITLE       Submitted (16-MAR-1999) Ravassard P., Lgn, CNRS UMR 9923, Hopital
            de la Pitie Salpêtrière, Bat. CERVI, 83 Bd. de l'Hopital, 75013
            PARIS, FRANCE
FEATURES
            Location/Qualifiers
            1..1330
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               1..1330
                  /gene="ngn3"
                  join(<1..157,321..>1330)
                  /gene="ngn3"
                  1..157
                     /gene="ngn3"
                     /number=1

```

```

intron      158..320
            /gene="ngn3"
            /number=1
exon        321..1330
            /gene="ngn3"
            /number=2
            322..366
            /gene="ngn3"
            /function="early neurogenesis"
            /codon_start=1
            /product="neurogenin 3"
            /protein_id="CAB45384.1"
            /db_xref="GI:5123783"
            /translation="MTQPSGAPTVQVTRTERSPPRASEDEVTCPTSPSPTRTPG
            NCAEAEBCRGAPRKLARRGRSPKSEIALSKORSRKANDRRNRMDLNSA
            LDALRGVPTFPDDAKLRIETLRFAHWIALQTLRIADHSLYALEPPAPHCGLG
            SPGRPRGMSLVSYPVQAGSLSPAISIEERGLIGATSSACLSRGLAFSDFL"
BASE COUNT    230 a      459 c      413 g      228 t
ORIGIN

```

```

Query Match      2.2%; Score 32; DB 9; Length 1330;
Best Local Similarity 100.0%; Pred. No. 1.4e-06;
Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      672 AAGACGAGTTGGCACTGACGACGCGACG 703
Db      535 AAGACGAGTTGGCACTGACGACGCGACG 566

```

```

Search completed: January 29, 2004, 21:29:57
Job time : 5314 secs

```


Query Match 6.3%; Score 92; DB 4; Length 804;
Best Local Similarity 100.0%; Pred. No. 9.9e-37;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GCGGTGATGGCTGGCGGCTCTCTGCCACCTTCCGGATGAGCCCAACTTCAAG 821
DB 463 GCGGTGATGGCTGGCGGCTCTCTGCCACCTTCCGGATGAGCCCAACTTCAAG 522
OY 822 ATCGAGACCTGGCGCTTGGCCCAACTACAT 853
DB 523 ATCGAGACCTGGCGCTTGGCCCAACTACAT 554

RESULT 2

US-08-722-570-13
Sequence 13, Application US/08722570
Patent No. 6555337
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Oifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/722,570
FILING DATE: 27-SEP-1996
CLASSIFICATION: 5365
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.
REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 738 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
US-08-722-570-13

Query Match 1.6%; Score 23; DB 4; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754
DB 304 GAGCGCAACCGCATGCACACCT 326

RESULT 3

US-08-932-411A-13
Sequence 13, Application US/08932411A
Patent No. 6566496
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Oifu
TITLE OF INVENTION: NEUROGENIN

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996

ATTORNEY/AGENT INFORMATION:

NAME: Silva, Robin M.

REGISTRATION NUMBER: 38,304

REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 738 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: CDS

LOCATION: 1..732

US-08-932-411A-13

Query Match 1.6%; Score 23; DB 4; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754
DB 304 GAGCGCAACCGCATGCACACCT 326

RESULT 4

US-08-910-973-21
Sequence 21, Application US/08910973
Patent No. 5795723
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910.973
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239.238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FPCR-1-10958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Mus musculus
IMMEDIATE SOURCE:
CLONE: neuroD3
FEATURE:
NAME/KEY: CDS
LOCATION: 101..835
US-09-499-227-21

```

```

Query Match 1.6%; Score 23; DB 1; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 732 GAGCGCAACCGCATGCACAACT 754
Db 404 GAGCGCAACCGCATGCACAACT 426

RESULT 5
US-09-499-227-21
Sequence 21, Application US/09499227
Patent No. 6444463
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/499,227
FILING DATE: 05-August-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-May-1994
PRIOR APPLICATION DATA:

```

```

APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-May-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/910.973
FILING DATE: 07-August-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FPCR-1-12742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 1333 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Mus musculus
IMMEDIATE SOURCE:
CLONE: neuroD3
FEATURE:
NAME/KEY: CDS
LOCATION: 101..835
US-09-499-227-21

```

```

Query Match 1.6%; Score 23; DB 4; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 732 GAGCGCAACCGCATGCACAACT 754
Db 404 GAGCGCAACCGCATGCACAACT 426

RESULT 6
US-08-932-411A-17
Sequence 17, Application US/08932411A
Patent No. 6566496
GENERAL INFORMATION:
APPLICANT: Anderson, David J.
APPLICANT: Ma, Qifu
TITLE OF INVENTION: NEUROGENIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/932,411A
FILING DATE: 15-SEP-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/772,009
FILING DATE: 19-DEC-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/722,570
FILING DATE: 19-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Silva, Robin M.

```

REGISTRATION NUMBER: 38,304
REFERENCE/DOCKET NUMBER: A-63902-3/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 1385 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 382..1170
US-08-932-411A-17

Query Match 1.6%; Score 23; DB 4; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.072;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 732 GAGCGCAACCGCATGCACACCT 754
DB 742 GAGCGCAACCGCATGCACACCT 764

RESULT 7
US-08-552-142A-12
Sequence 12, Application US/08552142A
Patent No. 5695995
GENERAL INFORMATION:
APPLICANT: Weintraub, Harold M.
APPLICANT: Lee, Jacqueline E.
APPLICANT: Tapscott, Stephen J.
APPLICANT: Hollenberg, Stanley M.
TITLE OF INVENTION: Neurogenic Differentiation (Neurod) Genes
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/552,142A
FILING DATE: 02-NOV-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Broderick, Thomas P.
REGISTRATION NUMBER: 31,332
REFERENCE/DOCKET NUMBER: FHCR-1-8933
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100
TELEFAX: 206-225-0709
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 310 base pairs
TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: 20A1
FEATURE:
NAME/KEY: CDS
LOCATION: 1..310
US-08-552-142A-12

Query Match 1.4%; Score 21; DB 1; Length 310;
Best Local Similarity 100.0%; Pred. No. 0.72;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

CY 732 GAGCGCAACCGCATGCACAC 752
DB 128 GAGCGCAACCGCATGCACAC 148

RESULT 8
US-08-910-973-12
Sequence 12, Application US/08910973
Patent No. 5795723
GENERAL INFORMATION:
APPLICANT: Tapscott, Stephen J.
APPLICANT: Olson, James M.
TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen O'Connor Johnson Kindness PLLC
STREET: 1420 Fifth Avenue, Suite 2800
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101-2347
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,973
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,238
FILING DATE: 06-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US95/05741
FILING DATE: 08-MAY-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/17532
FILING DATE: 30-October-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sheiness, Diana K.
REGISTRATION NUMBER: 35,356
REFERENCE/DOCKET NUMBER: FHCR-1-10958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-682-8100; 206-224-0735 (direct)
TELEFAX: 206-225-0779
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 1268 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: 20A1 (neuroD3)
FEATURE:

NAME/KEY: CDS
LOCATION: 55..768
US-08-910-973-12

Query Match 1.4%; Score 21; DB 1; Length 1268;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752
DB 355 GAGCGCAACCGCATGCACAC 375

RESULT 9
US-09-499-227-12

Sequence 12, Application US/09499227
Patent No. 6444463

GENERAL INFORMATION:

APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.

TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen O'Connor Johnson Kindness PLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/499,227

FILING DATE: 05-August-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-May-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-May-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US96/17532

FILING DATE: 30-October-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/910,973

FILING DATE: 07-August-1997

ATTORNEY/AGENT INFORMATION:

NAME: Sheiness, Diana K.

REGISTRATION NUMBER: 35,356

REFERENCE/DOCKET NUMBER: FHC-1-12742

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100; 206-224-0735 (direct)

TELEFAX: 206-225-0779

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 1268 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: 20A1 (neuroD3)

FEATURE:

NAME/KEY: CDS

LOCATION: 55..768

US-09-499-227-12

Query Match 1.4%; Score 21; DB 4; Length 1268;

Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752
DB 355 GAGCGCAACCGCATGCACAC 375

RESULT 10
US-08-552-142A-10

Sequence 10, Application US/08552142A
Patent No. 5695995

GENERAL INFORMATION:

APPLICANT: Weintraub, Harold M.

APPLICANT: Lee, Jacqueline E.

APPLICANT: Tapscott, Stephen J.

TITLE OF INVENTION: Neurogenic Differentiation (NeuroD) Genes

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESS: Christensen O'Connor Johnson Kindness PLLC

STREET: 1420 Fifth Avenue, Suite 2800

CITY: Seattle

STATE: WA

COUNTRY: USA

ZIP: 98101-2347

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/552,142A

FILING DATE: 02-NOV-1995

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,238

FILING DATE: 06-May-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/US95/05741

FILING DATE: 08-May-1995

ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.

REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: FHC-1-8933

TELECOMMUNICATION INFORMATION:

TELEPHONE: 206-682-8100

TELEFAX: 206-225-0709

INFORMATION FOR SEQ ID NO: 10:

SEQUENCE CHARACTERISTICS:

LENGTH: 1352 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: 14B1

FEATURE:

NAME/KEY: CDS

LOCATION: 55..1194

US-08-552-142A-10

Query Match 1.4%; Score 21; DB 1; Length 1352;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 729 CGGAGCGCAACCGCATGCAC 749
DB 439 CGGAGCGCAACCGCATGCAC 459

RESULT 11
US-08-910-973-10
; Sequence 10, Application US/08910973
; Patent No. 5795723
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.
; APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson KindnessAPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,973
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-10958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 14B1 (neuroD2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..1194
US-08-910-973-10

Query Match 1.4%; Score 21; DB 1; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CCGGAGCGCAACCGCATGCAC 749
DB 439 CCGGAGCGCAACCGCATGCAC 459

RESULT 12
US-09-499-227-10
; Sequence 10, Application US/09499227
; Patent No. 6444463
; GENERAL INFORMATION:
; APPLICANT: Tapscott, Stephen J.

APPLICANT: Olson, James M.
; TITLE OF INVENTION: Expression of Neurogenic bHLH Genes in Primitive Neuroectoder
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen O'Connor Johnson KindnessAPLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101-2347
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/499,227
; FILING DATE: 05-August-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/239,238
; FILING DATE: 06-MAY-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/US95/05741
; FILING DATE: 08-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/17532
; FILING DATE: 30-October-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/910,973
; FILING DATE: 07-August-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Sheiness, Diana K.
; REGISTRATION NUMBER: 35,356
; REFERENCE/DOCKET NUMBER: FHCR-1-12742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 206-682-8100; 206-224-0735 (direct)
; TELEFAX: 206-225-0779
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1535 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: 14B1 (neuroD2)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 55..1194
US-09-499-227-10

Query Match 1.4%; Score 21; DB 4; Length 1535;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 CCGGAGCGCAACCGCATGCAC 749
DB 439 CCGGAGCGCAACCGCATGCAC 459

RESULT 13
US-09-234-332-3
; Sequence 3, Application US/09234332A
; Patent No. 6087168
; GENERAL INFORMATION:
; APPLICANT: Cedars-Sinai Medical Center
; APPLICANT: Michael F. Levesque, M.D.
; APPLICANT: Thomas Neuman, Ph.D.
; TITLE OF INVENTION: CONVERSION OF NON-NEURONAL CELLS INTO
; TITLE OF INVENTION: NEURONS; TRANSDIFFERENTIATION OF EPIDERMAL CELLS
; FILE REFERENCE: P07 41494

CURRENT APPLICATION NUMBER: US/09/234,332A
CURRENT FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 16
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO: 3
LENGTH: 1550
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: gene
LOCATION: (0)...(0)
OTHER INFORMATION: Neurogenic basic-helix-loop-helix protein (Neuro
OTHER INFORMATION: D2) gene Genbank Accession U58681
FEATURE:
NAME/KEY: unsure
LOCATION: (1219)...(1226)
OTHER INFORMATION: n at 1219 and 1226; n = A, T, G, or C
US-09-234-332-3

Query Match 1.4%; Score 21; DB 3; Length 1550;
Best Local Similarity 100.0%; Pred. No. 0.74;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 729 CCGGAGCGCAACCGCATGCAC 749
Db 442 CCGGAGCGCAACCGCATGCAC 462

RESULT 14
US-08-358-627F-4/C
Sequence 4, Application US/08358627F
Patent No. 6177242
GENERAL INFORMATION:
APPLICANT: Changeux, Jean-Pierre
APPLICANT: Picciotto, Marina
TITLE OF INVENTION: Genomic DNA Fragments Containing
TITLE OF INVENTION: Regulatory and Coding Sequences For The R2-Subunit of the
TITLE OF INVENTION: Neuronal Nicotinic Acetylcholine Receptor and Transgenic
TITLE OF INVENTION: Animals Made Using These Fragments or Mutated Fragments
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Flinnegan, Henderson, Farabow, Garrett &
ADDRESSEE: Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/358,627F
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Meyers, Kenneth J.
REGISTRATION NUMBER: 25,146
REFERENCE/DOCKET NUMBER: 03495.0135-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4400
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-358-627F-4

Query Match 1.3%; Score 19; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAGGTAGCGAGGAGCA 19
Db 33 GCAGGTAGCGAGGAGCA 15

RESULT 15
US-08-793-044-11/C
Sequence 11, Application US/08793044
Patent No. 6235497
GENERAL INFORMATION:
APPLICANT: Benjamin, Stephane
APPLICANT: Berrard, Sylvie
APPLICANT: Cervini, Riccardo
APPLICANT: Maillet, Jacques
TITLE OF INVENTION: NOVEL VESICULAR ACETYLCHOLINE CARRIER
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Rhone-Poulenc Rorer Inc.
STREET: 500 Arcola Road, Mailstop 3C43
CITY: Collegeville
STATE: PA
COUNTRY: USA
ZIP: 19426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/793,044
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/FR95/01073
FILING DATE: 10-AUG-1995
PRIOR APPLICATION DATA: FR 94/10044
APPLICATION NUMBER: FR 94/10044
FILING DATE: 16-AUG-1994
ATTORNEY/AGENT INFORMATION:
NAME: Savitzky Esq., Martin F.
REGISTRATION NUMBER: 29,699
REFERENCE/DOCKET NUMBER: ST94066-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (610) 454-3816
TELEFAX: (610) 454-3808
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 50 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-793-044-11

Query Match 1.3%; Score 19; DB 3; Length 50;
Best Local Similarity 100.0%; Pred. No. 7.2;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GCAGGTAGCGAGGAGCA 19
Db 33 GCAGGTAGCGAGGAGCA 15

Search completed: January 29, 2004, 22:20:35
Job time : 119 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      January 26, 2004, 19:33:06 ; Search time 426 Seconds
              (without alignments)
              9251.595 Million cell updates/sec
```

Title: US-09-595-947E-1

Sequence: 1 gcaggtacgagagagagcag.....agagtgcactaatccagtgt 1460

Scoring table: OLIGO_NUC

Searched: 2552756 seqs, 1349719017 residues

```
word size : 0
```

Total number of hits satisfying chosen parameters: 5105512

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Listing first 45 summaries

Database : N_Geneseq_19Jun03:*

1	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001.DAT.*
23	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25	/SIDS1/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	1409	96.5	1491	19	AAV42512	CDNA encoding a no
2	92	6.3	804	21	AAV27050	Mouse neurogenin 3
3	92	6.3	804	21	AAZ51981	Mouse neurogenin 3
4	92	6.3	861	22	AAZ27266	Mouse neurogenin 3
5	92	6.3	861	25	ABY75570	Mouse transcriptio
6	92	6.3	1860	24	AAV46872	Mouse neurogenin
7	92	6.3	1861	21	AAAC61090	Mouse neurogenin
8	92	6.3	5567	22	AAV27254	Mouse atonal homol

9	36	2.5	65	24	ABN13392
10	35	2.4	65	24	ABN57521
11	35	2.4	6123	24	AAD668989
12	32	2.2	5340	21	AAC10889
13	32	2.2	5340	21	AAD668971
14	29	2.0	428	22	AAS337007
15	26	1.8	26	24	ABT037909
16	25	1.7	25	19	AAV42515
17	23	1.6	738	19	AAV27044
18	23	1.6	738	21	AAZ519777
19	23	1.6	790	22	AAV272628
20	23	1.6	1332	19	AAV42938
21	23	1.6	1332	25	ABSS63999
22	23	1.6	1332	18	AAV248894
23	23	1.6	1385	19	AAV270434
24	23	1.6	1385	21	AAZ51980
25	23	1.6	1385	22	AAV272659
26	23	1.6	1412	22	AAV272525
27	23	1.6	1412	22	AAV272273
28	21	1.4	21	25	AAD47278
29	21	1.4	21	25	AAD47278
30	21	1.4	714	24	AAD668888
31	21	1.4	1268	18	AAV448931
32	21	1.4	1268	19	AAV42932
33	21	1.4	1268	25	ABSS63999
34	21	1.4	1535	18	AAV48990
35	21	1.4	1535	19	AAV42931
36	21	1.4	1535	25	ABSS63889
37	21	1.4	1550	21	AAV626818
38	21	1.4	1665	24	AAD668898
39	21	1.4	2776	22	AAV40404
40	21	1.4	2776	22	AAV40404
41	21	1.4	2776	22	AAK68475
42	21	1.4	2776	22	AAK68476
43	20	1.4	352	24	ABSS90404
44	20	1.4	592	24	ABQ495522
45	20	1.4	592	24	ABQ495522

ALIGNMENTS

RESULT 1	
AAV42512	
ID	AAV42512 standard; cDNA; 1491 BP.

AC AAV42512;

DE CDNA encoding a novel BHLH protein designated RELAX

KW Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;

KW protein expression; central nervous system; CNS; treatment;

30

100-443887-1000

CDS 459.:1103
ET /*aa- 3
ET

```

      /product= RELAY
      FT
      XX

```

MO962/206-AZ
FN
XX

FD-302a (Rev. 11-27-83)

XX

XX

XX

FT /product= "Mouse neurogenin 3"

XX WO9813491-A2.

XX 02-APR-1998.

XX 24-SEP-1997; 97WO-US17048.

XX 17-SEP-1997; 97US-0932411.

XX 27-SEP-1996; 96US-0722570.

XX 12-NOV-1996; 96US-0030864.

XX 19-DEC-1996; 96US-0772009.

XX (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Ma Q, Sommer L;

XX WPI; 1998-230702/20.

XX P-PSDB; AAM54947.

XX Mouse neurogenin, useful in neurogenesis - and recombinant nucleic

XX acids and proteins derived from rat and Xenopus

XX Disclousure; Fig 9, 106pp; English.

XX The Mouse neurogenin 3 is one of several neurogenin proteins discussed

XX in the present invention. The neurogenin nucleic acids can be expressed

XX in a host cell, transformed using an expression vector, to produce

XX recombinant proteins. The proteins and the antibodies raised against

XX the proteins are useful in the study of neurogenesis.

XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

XX

XX Query Match 6.3%; Score 92; DB 19; Length 804;

XX Best Local Similarity 100.0%; Pred. No. 9e-35;

XX Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 762 GCGCTGATGCGCTGCGGCTGCTGCGCCACCTTCCCGATGAGCCCAACTTACAAG 821

XX 463 GCGCTGATGCGCTGCGGCTGCTGCGCCACCTTCCCGATGAGCCCAACTTACAAG 522

XX

XX 822 ATCGAGACCTTGCGCTTGCGCCCAACTTACAT 853

XX 523 ATCGAGACCTTGCGCTTGCGCCCAACTTACAT 554

XX

XX RESULT 3

XX AA251981

XX ID AA251981 standard; DNA; 804 BP.

XX AA251981;

XX 04-JUN-2000 (first entry)

XX Murine neurogenin-3 (NGN3) nucleic acid sequence.

XX

XX Neurogenin-3; NGN-3; non-neuronal cell; NNC; neurogenesis;

XX Phox2a protein; neuronal subtype-specific marker; growth factor;

XX neural differentiation; transplantation; neuronal dysfunction;

XX optical nerve damage; auditory nerve damage; neurodegenerative disorder;

XX neuroprotective; neurotropic; anticonvulsant; antiParkinsonian; vulnerary;

XX cerebroprotective; immunosuppressant; antiinfectious; ds.

XX Mus sp.

XX Key Location/Qualifiers

XX CDS 160..804

XX /tag= a

XX /product= "Murine neurogenin-3 protein"

XX WO200009676-A2.

XX 24-FEB-2000.

XX 13-AUG-1999; 99WO-US18525.

XX 14-AUG-1998; 98US-0096630.

XX (CALY) CALIFORNIA INST OF TECHNOLOGY.

XX Anderson DJ, Lo L;

XX WPI; 2000-256250/22.

XX P-PSDB; AAY70570.

XX Inducing non-neuronal cells to differentiate into neurons and for

XX non-neuronal cells to express a neuronal subtype-specific marker,

XX comprising contacting the non-neuronal cells with a vector containing

XX neurogenin nucleic acid -

XX Claim 1; Fig 1J; 76pp; English.

XX The patent discloses a method for inducing non-neuronal cells (NNC) to

XX differentiate into neurons and for NNCs to express a neuronal subtype

XX -specific marker. Transformed host cells are used as sources of neuronal

XX and other growth factors, in culture for screening compounds that

XX modulate neural differentiation or as sources of recombinantly produced

XX neurogenins and Phox2a proteins for use in transplantation. The cells

XX also have a variety of in vivo uses, e.g. for transplantation at sites of

XX neuronal dysfunction e.g. patients with hearing or vision loss due to

XX optical or auditory nerve damage, brain or spinal cord injuries, and

XX neurodegenerative disorders e.g. Alzheimer's disease. The present

XX sequence encodes murine neurogenin-3 (NGN-3), a transcription factor.

XX NNCs differentiate into neurons through the recombinant expression of a

XX transcription factor that induces a core program of neurogenesis. Forced

XX expression of murine NGN3 can elicit expression of at least some neuronal

XX phenotypic markers even in NNCs.

XX Sequence 804 BP; 171 A; 263 C; 225 G; 145 T; 0 other;

XX

XX Query Match 6.3%; Score 92; DB 21; Length 804;

XX Best Local Similarity 100.0%; Pred. No. 9e-35;

XX Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX

XX 762 GCGCTGATGCGCTGCGGCTGCTGCGCCACCTTCCCGATGAGCCCAACTTACAAG 821

XX 463 GCGCTGATGCGCTGCGGCTGCTGCGCCACCTTCCCGATGAGCCCAACTTACAAG 522

XX

XX 822 ATCGAGACCTTGCGCTTGCGCCCAACTTACAT 853

XX 523 ATCGAGACCTTGCGCTTGCGCCCAACTTACAT 554

XX

XX RESULT 4

XX AAF27266

XX ID AAF27266 standard; cDNA; 861 BP.

XX AAF27266;

XX 24-APR-2001 (first entry)

XX Mouse neurogenin 3 (ngn3) cDNA, SEQ ID NO:24.

XX

XX Atonal; homologue; orthologue; atonal-associated protein; deafness;

XX hearing impairment; vestibular effect; balance disorder; osteoarthritis;

XX cellular proliferation; cerebellar granule neuron; gene therapy;

XX mechanoreceptive cell growth; auditory; osteopathic; cyostatic;

XX transgenic animal; ss.

XX Mus musculus.

XX WO2000073764-A2.

XX 07-DEC-2000.

XX 01-JUN-2000; 2000WO-US15410.

XX 01-JUN-1999; 99US-0137060.
 PR 19-JAN-2000; 2000US-0176993.
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 XX Zoghbi HY, Ballen H, Birmingham N, Haasan B, Ben-Arie N;
 PI MPI; 2001-032190/04.
 DR P-PSDB; AAB60359.
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness,
 PT osteoarthritis and abnormal cell proliferation -
 XX
 PS Disclosure; Page -; 142pp; English.
 XX The invention relates to the use of atonal-associated nucleic acid or
 CC amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 XX
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;
 Query Match 6.3%; Score 92; DB 22; Length 861;
 Best Local Similarity 100.0%; Pred. No. 9e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GGGCTGATGGCGCTGGCGGTCTCTGCGCACTTCCGGATGAGCGCAAACTTACAAG 821
 DB 463 GGGCTGATGGCGCTGGCGGTCTCTGCGCACTTCCGGATGAGCGCAAACTTACAAG 522
 QY 822 ATCGAGACCTGCGCTTGGCCCAACTACAT 853
 DB 523 ATCGAGACCTGCGCTTGGCCCAACTACAT 554
 RESULT 5
 ABV75970 standard; cDNA; 861 BP.
 XX ABV75970;
 XX 11-FEB-2003 (first entry)
 DE Mouse transcription factor neurogenin 3 cDNA.
 XX Mouse; transcription factor; neurogenin 3; ngn3; stem cell;
 KW differentiation; beta-cell; insulin; diabetes; hyperglycaemia;
 KW glucose intolerance; antidiabetic; hypoglycaemic; gene therapy;
 KW gene; ss.
 XX Mus musculus.
 OS
 FH Key Location/Qualifiers
 FT CDS 160..804
 FT /*tag= a

FT /product= "Mouse ngn3"
 XX WO200286107-A2.
 XX 31-OCT-2002.
 XX 19-APR-2002; 2002WO-EP04362.
 XX 19-APR-2001; 2001US-284531P.
 XX (DEVE-) DEVELOPEN ENTWICKLUNGSBIOLOGISCHE FORSCH.
 XX (PFLA-) INST PFLANZENGENETIK & KULTURPFLANZENFOR.
 XX Wobus AM, St-Onge L, Blyszczuk P, Hoffmann U;
 PI MPI; 2003-075629/07.
 XX
 DR Differentiating stem cells into insulin-producing cells useful for
 PT treating pancreatic diseases, by culturing stem cells in suitable
 PT medium and activating gene involved in beta-cell differentiation -
 XX
 PS Disclosure; Page 58-59; 62pp; English.
 XX The present sequence is that of cDNA encoding the murine
 CC basic helix-loop-helix transcription factor neurogenin 3 (ngn3),
 CC a gene which is required for the specification of the early
 CC endocrine precursor in the pancreatic epithelium and which is
 CC down-regulated once endocrine differentiation begins. The invention
 CC provides a claimed method for differentiating stem cells (especially
 CC embryonic, adult or somatic stem cells and primordial germ cells)
 CC into insulin-producing cells. This involves culturing stem cells in
 CC a suitable medium and activating at least one gene involved in
 CC beta-cell differentiation. Preferred genes including Pdx1, Pax4,
 CC Pax6 and ngn3 (see ABV75967-70). Gene activation comprises the
 CC delivery of the gene into stem cells using a viral delivery
 CC system, or the delivery of a protein product of the gene into stem
 CC cells. The insulin-producing cells can be transplanted into
 CC animals or human for treatment of pancreatic diseases, metabolic
 CC syndrome and metabolic disorders with impaired glucose levels such
 CC as diabetes, hyperglycaemia and impaired glucose tolerance
 CC (claimed). The cells can also be used to identify compounds which
 CC stimulate beta-cell differentiation, insulin secretion or glucose
 CC responsiveness. Differentiated beta-cells can be used to study the
 CC toxic and other effects of exogenous compounds on beta-cell
 CC function. In an example from the invention, Pax6 cDNA was inserted
 CC into expression vector pACCMV.plpa under the control of the
 CC cytomegalovirus promoter.
 XX
 SQ Sequence 861 BP; 182 A; 274 C; 250 G; 155 T; 0 other;
 Query Match 6.3%; Score 92; DB 25; Length 861;
 Best Local Similarity 100.0%; Pred. No. 9e-35;
 Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 762 GGGCTGATGGCGCTGGCGGTCTCTGCGCACTTCCGGATGAGCGCAAACTTACAAG 821
 DB 463 GGGCTGATGGCGCTGGCGGTCTCTGCGCACTTCCGGATGAGCGCAAACTTACAAG 522
 QY 822 ATCGAGACCTGCGCTTGGCCCAACTACAT 853
 DB 523 ATCGAGACCTGCGCTTGGCCCAACTACAT 554
 RESULT 6
 AAD46872 standard; DNA; 1860 BP.
 XX AAD46872;
 XX 27-JAN-2003 (first entry)
 DT Murine neurogenin 3 (Ngn3) gene.
 XX

Transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
islet cell; cell therapy; neurogenin 3; Ngn3; murine; gene; ds.
Mus musculus.
Key Location/Qualifiers
CDS 1093..1737
FT /tag= a
FT /product= "Murine Ngn3 protein"
WO200274045-A2.
26-SEP-2002.
20-MAR-2002; 2002WO-US11166.
20-MAR-2001; 2001US-0817360.
(REGC) UNIV CALIFORNIA.
German MS, Lin J;
WPI; 2002-759853/82.
P-PSDB; AAE29278.
Producing a mammalian islet cell for treating diabetes mellitus
comprises introducing into a mammalian cell a nucleic acid molecule
encoding neuroendocrine basic helix-loop-helix transcription factor -
Example 3; Page 89-90; 108pp; English.
The invention relates to a method for producing a mammalian islet cell.
The method comprising introducing into a mammalian cell a nucleic acid
molecule encoding an islet transcription factor for expression of the
islet transcription factor in the cell and for production of islet cell
phenotype in the cell. The islet transcription factor is a neuroendocrine
basic helix-loop-helix (bHLH) transcription factor. The method is useful
for treating type 2 diabetes mellitus and for replacing beta cells lost
to autoimmune destruction in individuals with type 1 diabetes. The method
is useful in cell therapy. The present sequence is murine neurogenin 3
(Ngn3) gene.
Sequence 1860 BP; 397 A; 559 C; 537 G; 367 T; 0 other;
Query Match 6.3%; Score 92; DB 24; Length 1860;
Best Local Similarity 100.0%; Pred. No. 8.7e-35;
Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 762 GGGCTGATGCGCTGCGGATGCTGCGCACCTTCGCGATGAGCGCAACTTACAAG 821
DB 1396 GGGCTGATGCGCTGCGGATGCTGCGCACCTTCGCGATGAGCGCAACTTACAAG 1455
QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
DB 1456 ATCGAGACCTGCGCTTCGCCCAACTACAT 1487
RESULT 7
AAC61090
ID AAC61090 standard; DNA; 1861 BP.
XX
AC AAC61090;
XX
DT 05-FEB-2001 (first entry)
XX
DE Murine neurogenin 3 (Ngn3) genomic DNA sequence.
XX
KW Neurogenin 3; Ngn3; cellular differentiation; diabetes mellitus;
KW islet cell precursor identification; mouse; ds.
XX
OS Mus musculus.
XX

Key Location/Qualifiers
CDS 1093..1737
FT /tag= a
FT /product= "Ngn3"
FT /note= "Neurogenin 3"
WO200059936-A1.
12-OCT-2000.
28-MAR-2000; 2000WO-US08436.
06-APR-1999; 99US-0128180.
(REGC) UNIV CALIFORNIA.
German MS, Lin J;
WPI; 2000-664989/64.
P-PSDB; AAY85618.
Novel human neurogenin 3 polypeptides and polynucleotides encoding
them, useful for diagnosis, prevention and treatment of diabetes
mellitus and to identify individuals at risk of diabetes -
Claim 18; Page 49-50; 54pp; English.
The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
protein AAY85617. The Ngn3 gene is located at chromosome position
10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC protein sequences, and includes an antibody recognising the Ngn3 protein.
CC Also included in the invention is a method for identifying an islet cell
CC precursor, the method involves analysing a cell for the expression of the
CC Ngn3 gene product, where detection of the product is indicative of an
CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC reagent for detecting (in a subject) a predisposition to a defect in
CC pancreatic islet cell function or formation associated with a defect in
CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC precursor cells expressing Ngn3, and to alter cellular differentiation in
CC culture in vivo to produce new beta-cells to treat patients with diabetes
CC mellitus. The present sequence represents the murine Ngn3 genomic DNA
CC sequence.
Sequence 1861 BP; 397 A; 560 C; 537 G; 367 T; 0 other;
QY 762 GGGCTGATGCGCTGCGGATGCTGCGCACCTTCGCGATGAGCGCAACTTACAAG 821
DB 1396 GGGCTGATGCGCTGCGGATGCTGCGCACCTTCGCGATGAGCGCAACTTACAAG 1455
QY 822 ATCGAGACCTGCGCTTCGCCCAACTACAT 853
DB 1456 ATCGAGACCTGCGCTTCGCCCAACTACAT 1487
RESULT 8
AAF27254
ID AAF27254 standard; cDNA; 5567 BP.
XX
AC AAF27254;
XX
DT 24-APR-2001 (first entry)
XX
DE Mouse atonal homologue 5 (ATOH5, Math4B) cDNA, SEQ ID NO:4.
XX
KW Atonal; homologue; orthologue; atonal-associated protein; deafness;
KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KW cellular proliferation; cerebellar granule neuron; gene therapy;
KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
KW transgenic animal; ss.
KW

```

XX OS Mus musculus.
XX PN WO200073764-A2.
XX PD 07-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US15410.
XX PR 01-JUN-1999; 99US-0137060.
XX PR 19-JAN-2000; 2000US-0176993.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PI Zoghbi HY, Bellien H, Birmingham N, Haassan B, Ben-Arie N;
XX DR WPI, 2001-032190/04.
XX DR P-PSDB; AAB60350.
XX PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
XX PT any of its homologs or orthologs, for the treatment of e.g. deafness,
XX PT osteoarthritis and abnormal cell proliferation -
XX PS Disclosure; Page -: 142pp; English.
XX CC The invention relates to the use of atonal-associated nucleic acid or
XX CC amino acid sequence, or any of its homologues or orthologues as
XX CC therapeutic agents for the treatment of deafness, partial hearing loss,
XX CC vestibular effects due to damage or loss of inner hair cells,
XX CC osteoarthritis and abnormal cell proliferation. The invention also
XX CC encompasses methods of screening for compounds which affect the
XX CC expression of an atonal-associated nucleic acid sequence in an animal,
XX CC and a transgenic animal in which an allele of a native atonal-associated
XX CC gene is replaced by a heterologous nucleic acid sequence, thus
XX CC inactivating the atonal-associated allele. The nucleic acids or proteins
XX CC may be used in a method of treating an animal for hearing impairment,
XX CC joint disease, balance disorders, abnormal cell proliferation, or other
XX CC disease related to loss of a functional atonal-associated nucleic acid or
XX CC protein. They may particularly be used to treat an animal with a
XX CC deficiency in cerebellar granule neurons or their precursors, and may
XX CC also be used in promoting mechanoreceptive cell growth and generating
XX CC hair cells. The present sequence represents an atonal-associated nucleic
XX CC acid sequence referred to in the invention.
XX CC Note: The present sequence is not shown in the specification, but
XX CC was obtained from Genbank.
XX SQ Sequence 5567 BP; 1271 A; 1549 C; 1564 G; 1183 T; 0 other;
XX
XX Query Match 6.3%; Score 92; DB 22; Length 5567;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-35;
XX Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 762 GCGCTGATGCGCTGCGCGGTCTCTGCGCACCTTCCGGATGAGCGCAACTTACAAAG 821
DB 5226 GCGCTGATGCGCTGCGCGGTCTCTGCGCACCTTCCGGATGAGCGCAACTTACAAAG 5285
OY 822 ATCGAGACCCCTGCGCTTGCGCCCAACTACAT 853
DB 5286 ATCGAGACCCCTGCGCTTGCGCCCAACTACAT 5317

```

```

XX OS Rattus norvegicus.
XX PN WO200210449-A2.
XX PD 07-FEB-2002.
XX PF 20-JUL-2001; 2001WO-1B01903.
XX PR 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Fatgler S;
XX DR WPI, 2002-257383/30.
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes -
XX PS Example 1; SEQ ID 4140; 47pp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridizing selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterizing the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome of a patient suffering
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 65 BP; 11 A; 20 C; 14 G; 20 T; 0 other;
XX
XX Query Match 2.5%; Score 36; DB 24; Length 65;
XX Best Local Similarity 100.0%; Pred. No. 3.2e-07;
XX Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1281 ATTGAGGCTATCTCCTTAACCCCTCCTAGTGT 1316
DB 30 ATTGAGGCTATCTCCTTAACCCCTCCTCAGTGT 65

```

```

RESULT 9
ABN31392
ID ABN31392 standard; DNA; 65 BP.
XX
XX AC ABN31392;
XX
XX DT 15-JUL-2002 (first entry)
XX
XX DE Rat spliced transcript detection oligonucleotide SEQ ID NO:4140.
XX
XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

```

```

RESULT 10
ABN57521
ID ABN57521 standard; DNA; 65 BP.
XX
XX AC ABN57521;
XX
XX DT 15-JUL-2002 (first entry)
XX
XX DE Mouse spliced transcript detection oligonucleotide SEQ ID NO:30269.
XX
XX KW Human; mouse; rat; splice transcript; detection; RNA transcript;
XX splice variant; transcriptome; oligonucleotide library; ss.

```



```
XX OS Mus musculus.
XX OS WO200210449-A2.
XX PN 07-FEB-2002.
XX PD 20-JUL-2001; 2001WO-IB01903.
XX PF 28-JUL-2000; 2000US-221607P.
XX PR 02-MAY-2001; 2001US-287724P.
XX PA (COMP-) COMPUGEN INC.
XX PI Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;
XX WI WI: 2002-257383/30.
XX DR
XX PT New oligonucleotide libraries comprising oligonucleotides which
XX PT selectively hybridize to mRNAs transcribed from a transcription unit of
XX PT a genome, useful for detecting tissue-, pathology-, and
XX PT developmental-specific genes
XX PS
XX PS Example 1; SEQ ID 30269; 47bp; English.
XX CC The present invention describes oligonucleotide libraries for detecting
XX CC messenger RNAs that populate a (sub-)transcriptome, where the
XX CC (sub-)transcriptome comprises messenger RNAs transcribed from multiple
XX CC transcription units that populate a genome. The library comprises
XX CC several oligonucleotides, each capable of hybridizing selectively to a
XX CC set of messenger RNAs transcribed from a given transcription unit of
XX CC the genome, which encodes one or more messenger RNA splice variants.
XX CC The oligonucleotide libraries are useful for detecting mRNAs from a
XX CC biological sample, in expression profiling studies, in qualitatively or
XX CC quantitatively characterizing the corresponding transcriptome, and in
XX CC detecting RNA transcripts and splice variants of human or animal
XX CC transcriptomes. The libraries may also be used as specialised mini
XX CC libraries to detect transcripts of a sub-transcriptome under a
XX CC particular biological or pathological state, and so allowing the
XX CC detection of tissue- and pathology-specific genes such as those genes
XX CC only expressed in specific tissue under a specific pathological
XX CC condition; to detect developmental specific genes; and to detect RNA
XX CC transcripts and splice variants of a transcriptome; and to detect RNA
XX CC from a particular disorder. ABN27253 to ABN59589 represent
XX CC oligonucleotide sequences from rats, humans and mice, which are used in
XX CC the exemplification of the present invention.
XX CC N.B. The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 65 BP; 16 A; 25 C; 11 G; 13 T; 0 other;
XX
XX Query Match 2.4%; Score 35; DB 24; Length 65;
XX Best Local Similarity 100.0%; Pred. No. 9.8e-07;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 547 TCAGTTCGAATTCACCCCACTAGCCCACTCTC 581
XX DB 1 TCAGTTCGAATTCACCCCACTAGCCCACTCTC 35
XX
XX RESULT 11
XX ID AAD46890 standard; DNA; 6123 BP.
XX AC AAD46890;
XX XX 27-JAN-2003 (first entry)
XX DT Human neurogenin 2 (Ngn2) gene #1.
XX DE Human neurogenin 2 (Ngn2) gene #1.
XX KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
XX KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
```

```
KW islet cell; cell therapy; neurogenin 2; Ngn2; gene; ds.
XX OS Homo sapiens.
XX OS WO200274045-A2.
XX PN 26-SEP-2002.
XX PD 20-MAR-2002; 2002WO-US11166.
XX PF 20-MAR-2001; 2001US-0817360.
XX PR (REGC ) UNIV CALIFORNIA.
XX PA German MS, Lin J;
XX PI WI: 2002-759853/82.
XX DR P-PSDB; AAE29281.
XX XX
XX PT Producing a mammalian islet cell for treating diabetes mellitus
XX PT comprises introducing into a mammalian cell a nucleic acid molecule
XX PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX PS Disclosure; Page 95-97; 108bp; English.
XX CC The invention relates to a method for producing a mammalian islet cell.
XX CC The method comprising introducing into a mammalian cell a nucleic acid
XX CC molecule encoding an islet transcription factor for expression of the
XX CC islet transcription factor in the cell and for production of islet cell
XX CC phenotype in the cell. The islet transcription factor is a neuroendocrine
XX CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
XX CC for treating type 2 diabetes mellitus and for replacing beta cells lost
XX CC to autoimmune destruction in individuals with type 1 diabetes. The method
XX CC is useful in cell therapy. The present sequence is human neurogenin 2
XX CC (Ngn2) gene.
XX SQ Sequence 6123 BP; 1484 A; 1536 C; 1507 G; 1596 T; 0 other;
XX
XX Query Match 2.4%; Score 35; DB 24; Length 6123;
XX Best Local Similarity 100.0%; Pred. No. 8.1e-07;
XX Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 819 AAGATCGAGACCTGGCGTTGCCACACATCAT 853
XX DB 383 AAGATCGAGACCTGGCGTTGCCACACATCAT 417
XX
XX RESULT 12
XX ID AAC61089 standard; DNA; 5340 BP.
XX AC AAC61089;
XX XX 05-FEB-2001 (first entry)
XX DT Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX DE Human neurogenin 3 (Ngn3) genomic DNA sequence.
XX KW Neurogenin 3; Ngn3; chromosome 10q22.1-22.2; cellular differentiation;
XX KW islet cell precursor identification; diabetes mellitus; human; ds.
XX OS Homo sapiens.
XX XX
XX XX Key Location/Qualifiers
XX XX CDS 3022..3666
XX XX /*tag= a
XX XX /product= "Ngn3"
```

/note= "Neurogenin 3"

FT XX WO20005936-A1.
XX XX 12-OCT-2000.
XX PD 28-MAR-2000; 2000WO-US08436.
XX PF 06-APR-1999; 99US-0128180.
XX PR (REGC) UNIV CALIFORNIA.
XX PA German MS, Lin J;
XX PI WPI; 2000-664989/64.
XX DR P-PSDB; AAY85617.
XX DR Novel human neurogenin 3 polypeptides and polynucleotides encoding
PT them, useful for diagnosis, prevention and treatment of diabetes
PT mellitus and to identify individuals at risk of diabetes -
XX PS Claim 6; Page 46-48; 54pp; English.
XX CC The human neurogenin 3 Ngn3 DNA sequence AAC61089 encodes the Ngn3
CC protein AAY85617. The Ngn3 gene is located at chromosome position
CC 10q22.1-22.2. The invention relates to the human Ngn3 nucleotide and
CC protein sequences, and includes an antibody recognising the Ngn3 protein.
CC Also included in the invention is a method for identifying an islet cell
CC precursor, the method involves analysing a cell for the expression of the
CC Ngn3 gene product, where detection of the product is indicative of an
CC islet cell precursor. The Ngn3 DNA sequence is useful as a diagnostic
CC reagent for detecting (in a subject) a predisposition to a defect in
CC pancreatic islet cell function or formation associated with a defect in
CC Ngn3 activity. The Ngn3 protein is useful for identifying beta-cell
CC precursor cells expressing Ngn3, and to alter cellular differentiation in
CC culture in vivo to produce new beta-cells to treat patients with diabetes
CC mellitus.
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;
XX
XX Query Match 2.2%; Score 32; DB 21; Length 5340;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-05;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 672 AAGAGCGAGTTGGCACTGACGACGACG 703
DB 3235 AAGAGCGAGTTGGCACTGACGACGACG 3266

RESULT 13
ADD46871
ID ADD46871 standard; DNA; 5340 BP.
XX AC ADD46871;
XX DT 27-JAN-2003 (first entry)
XX DE Human neurogenin 3 (Ngn3) gene.
XX KW Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 3; Ngn3; chromosome 10q22.1-22.2;
KW gene; ds.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
FT CDS 3022..3666
FT /tag= a
FT /product= "Human Ngn3 protein"
XX PN WO200274045-A2.
XX

PD 26-SEP-2002.
XX PF 20-MAR-2002; 2002WO-US11166.
XX XX 20-MAR-2001; 2001US-0817360.
XX PR (REGC) UNIV CALIFORNIA.
XX PA German MS, Lin J;
XX PI WPI; 2002-759853/82.
XX DR P-PSDB; AAE29277.
XX DR Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX PS Example 2; Page 87-88; 108pp; English.
XX CC The invention relates to a method for producing a mammalian islet cell.
XX CC The method comprising introducing into a mammalian cell a nucleic acid
XX CC molecule encoding an islet transcription factor for expression of the
XX CC islet transcription factor in the cell and for production of islet cell
XX CC phenotype in the cell. The islet transcription factor is a neuroendocrine
XX CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
XX CC for treating type 2 diabetes mellitus and for replacing beta cells lost
XX CC to autoimmune destruction in individuals with type 1 diabetes. The method
XX CC is useful in cell therapy. The present sequence is human neurogenin 3
XX CC (Ngn3) gene. Ngn3 gene is located on chromosome 10q22.1-22.2.
XX SQ Sequence 5340 BP; 1215 A; 1500 C; 1514 G; 1111 T; 0 other;
XX
XX Query Match 2.2%; Score 32; DB 24; Length 5340;
XX Best Local Similarity 100.0%; Pred. No. 2.4e-05;
XX Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 672 AAGAGCGAGTTGGCACTGACGACGACG 703
DB 3235 AAGAGCGAGTTGGCACTGACGACGACG 3266

RESULT 14
AAS33797
ID AAS33797 standard; CDNA; 428 BP.
XX AC AAS33797;
XX XX 17-DEC-2001 (first entry)
XX DT Human CDNA encoding a novel foetal antigen, SEQ ID No 321.
XX DE Human; foetal tissue antigen; ss; antiinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytoskeletal; nephroprotective;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.
XX OS Homo sapiens.
XX PN WO200155312-A2.
XX PD 02-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01321.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX

PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-020515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 03-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0233397.
PR 14-SEP-2000; 2000US-0233398.
PR 14-SEP-2000; 2000US-0233399.
PR 14-SEP-2000; 2000US-0233400.
PR 14-SEP-2000; 2000US-0233401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249224.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254057.
PR 05-JAN-2001; 2001US-0259678.

(HUMA-) HUMAN GENOME SCI INC.
PA
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-488782/53.
DR P-PSDB; AAU20977.
XX
XX
PT New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems -
XX
XX
PS Claim 1; SEQ ID No 321, 642pp; English.
XX

CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
CC immunoassay assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 2.0%; Score 29; DB 22; Length 428;
Best Local Similarity 100.0%; Pred. No. 0.0008;
Matches 29; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 825 GAGACCTGCGCTGCGCCCAACAATACAT 853
Db 1 GAGACCTGCGCTGCGCCCAACAATACAT 29

RESULT 15
ABT03700/C
ID ABT03700 standard; DNA; 26 BP.

XX AC ABT03700;

XX DT 13-SEP-2002 (first entry)

XX DE Human Neurogenin-3 gene PCR primer SEQ ID NO: 221.

XX KM Human; cancer; neoplastic disease; tumour specific marker; cytosolatic;
XX KW transcription factor; PCR; primer; ss.

XX OS Homo sapiens.

XX PN WO200240716-A2.

XX PD 23-MAY-2002.

XX PF 13-NOV-2001; 2001WO-US43461.

XX PR 16-NOV-2000; 2000US-249508P.

XX PA (CEMI-) CEMINES LLC.

XX PI Palm K;

XX DR WPI; 2002-537346/57.

XX PT Determining the presence of neoplastic molecular markers, by
XX PT identifying the presence of markers in host test sample using array of
XX PT neoplastic molecular marker specific reagents and analyzing the array
XX PT of the reagents -

XX PS Example 1; Page 17; 41pp; English.

XX CC The present invention relates to a method for determining the presence of
XX CC neoplastic molecular markers in a host, involving the use of neoplastic
XX CC molecular marker specific reagents to detect such markers and analyzing

CC the array of reagents, allowing the identification of the neoplastic
CC disease present. This can be used to determine the best treatment for
CC cancer, in particular neural cell, lung and prostate tumours. The
CC present sequence is a PCR primer useful for detecting the coding
CC sequences of markers of the invention.

Qy Sequence 26 BP; 3 A; 10 C; 5 G; 8 T; 0 other;
Query Match 1.8%; Score 26; DB 24; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.027;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 673 AGAGCACTGCGCACTGAGCAAGCAG 698
Db 26 AGAGCACTGCGCACTGAGCAAGCAG 1

RESULT 16
AAV42515
ID AAV42515 standard; DNA; 25 BP.

XX AC AAV42515;

XX DT 05-OCT-1998 (first entry)

XX DE PCR primer used to isolate part of the RELAX protein coding region.

XX KM Basic helix-loop-helix; BHLH; RELAX; Rat Embryonic Longitudinal Axis;
XX KW control; gene expression; transcriptional activator; targeting;
XX KW protein expression; central nervous system; CNS; treatment;
XX KW nervous system disorder; CIG235; PCR primer; ss.

XX OS Synthetic.

XX OS Rattus sp.

XX PN WO9827206-A2.

XX PD 25-JUN-1998.

XX PF 19-DEC-1997; 97WO-FR02368.

XX PR 19-DEC-1996; 96FR-0015651.

XX PA (RHON) RHONE-POULENC RORER SA.

XX PI Mallet J, Ravassard P, Icard-Liepkalns C;

XX DR WPI; 1998-362775/31.

XX PT Basic helix-loop-helix polypeptide and related nucleic acid - with
XX PT transcriptional activity, for targeting expression of genes to
XX PT central nervous system and treatment of nervous disease

XX PS Example 2; Page 12; 28pp; French.

XX CC PCR primers AAV42515-16 are used to isolate part of the DNA encoding
XX CC a basic helix-loop-helix (BHLH) type protein, designated RELAX (Rat
XX CC Embryonic Longitudinal Axis) protein. The PCR product is termed CIG235.
XX CC The protein is used to control and participate in gene expression,
XX CC by acting as transcriptional activator, strictly dependent on the
XX CC presence of an intact E box (CANNTG), particularly for targeting
XX CC expression of proteins to the central nervous system (CNS). The
XX CC nucleic acid sequence can be used to treat nervous system disorders,
XX CC and antisense sequences can be used to control mRNA transcription.

XX SQ Sequence 25 BP; 5 A; 9 C; 6 G; 5 T; 0 other;

Query Match 1.7%; Score 25; DB 19; Length 25;
Best Local Similarity 100.0%; Pred. No. 0.083;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 750 AACCTTAATCGCGCTGATGCGC 774
|||||

Db 1 AACCTTAACCTCCGCGCTGATCGGC 25

RESULT 17
AAV27046
ID AAV27046 standard; CDNA, 738 BP.
XX
AC AAV27046;
XX
DT 17-SEP-1998 (first entry)
XX
DE Mouse neurogenin 1 gene.
XX
KW ss; Mouse; neurogenin; expression vector; recombinant protein;
antibody; neurogenesis.
XX
OS Mus sp.
XX
FH Key Location/Qualifiers
FT CDS 1..735
FT /*tag= a
FT /product= "Mouse neurogenin 1"
XX
XX MO9813491-A2.
XX
XX 02-APR-1998.
XX
XX 24-SEP-1997; 97WO-US17048.
XX
XX 17-SEP-1997; 97US-0932411.
XX 27-SEP-1996; 96US-0722570.
XX 12-NOV-1996; 96US-0030864.
XX 19-DEC-1996; 96US-0772009.
XX
XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Anderson DJ, Ma Q, Sommer L;
XX
XX MPI, 1998-230702/20.
XX
XX P-PSDB; AAW54944.
XX
XX Mouse neurogenin, useful in neurogenesis - and recombinant nucleic
PT acids and proteins derived from rat and Xenopus
XX
XX Claim 5; Fig 4; 106pp; English.
XX
XX The mouse neurogenin 1 is one of several neurogenin proteins discussed in
CC the present invention. The neurogenin nucleic acids can be expressed in
CC a host cell, transformed using an expression vector, to produce
CC recombinant proteins. The proteins and the antibodies raised against
CC the proteins are useful in the study of neurogenesis.
XX
XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;
SQ

Query Match 1.6%; Score 23; DB 19; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754
DB 304 GAGCGCAACCGCATGCACAACT 326

RESULT 18
AAZ51977
ID AAZ51977 standard; DNA; 738 BP.
XX
AC AAZ51977;
XX
DT 04-JUL-2000 (first entry)
XX
DE Murine neurogenin-1 (NGN1) nucleic acid sequence.
XX

KW Neurogenin-1; NGN-1; non-neuronal cell; NNC; neurogenesis;
KW Phox2a protein; neuronal subtype-specific marker; growth factor;
KW neural differentiation; transplantation; neuronal dysfunction;
KW optical nerve damage; auditory nerve damage; neurodegenerative disorder;
KW neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerrary;
KW cerebroprotective; immunosuppressant; antiinfectious; ss.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
FT CDS 1..735
FT /*tag= a
FT /product= "Murine neurogenin-1 protein"
XX
XX MO200009676-A2.
XX
XX 24-FEB-2000.
XX
XX 13-AUG-1999; 99WO-US18525.
XX
XX 14-AUG-1998; 98US-0096630.
XX
XX (CALY) CALIFORNIA INST OF TECHNOLOGY.
XX
XX Anderson DJ, Lo L;
XX
XX MPI; 2000-256250/22.
XX P-PSDB; AAY70566.
XX
XX Inducing non-neuronal cells to differentiate into neurons and for
PT non-neuronal cells to express a neuronal subtype-specific marker,
PT comprising contacting the non-neuronal cells with a vector containing
PT neurogenin nucleic acid -
XX
XX Claim 1; Fig 1C; 76pp; English.
XX
XX The patent discloses a method for inducing non-neuronal cells (NNC) to
CC differentiate into neurons and for NNCs to express a neuronal subtype
CC -specific marker. Transformed host cells are used as sources of neuronal
CC and other growth factors; in culture for screening compounds that
CC modulate neural differentiation or as sources of recombinantly produced
CC neurogenins and Phox2a proteins for use in transplantation. The cells
CC also have a variety of in vivo uses, e.g. for transplantation at sites of
CC neuronal dysfunction e.g. patients with hearing or vision loss due to
CC optical or auditory nerve damage, brain or spinal cord injuries, and
CC neurodegenerative disorders e.g. Alzheimer's disease. The present
CC sequence encodes murine neurogenin-1 (NGN-1), a transcription factor.
CC NNCs differentiate into neurons through the recombinant expression of a
CC transcription factor that induces a core program of neurogenesis. Forced
CC expression of murine NGN1 can elicit expression of at least some neuronal
CC phenotypic markers even in NNCs. This can be used in autografting.
XX
XX Sequence 738 BP; 121 A; 283 C; 205 G; 129 T; 0 other;
SQ

Query Match 1.6%; Score 23; DB 21; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.69;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACT 754
DB 304 GAGCGCAACCGCATGCACAACT 326

RESULT 19
AAF27264
ID AAF27264 standard; CDNA; 790 BP.
XX
AC AAF27264;
XX
DT 24-APR-2001 (first entry)
XX
DE Chicken atonal homologue ngn2/ath4a CDNA, SEQ ID NO:20.
XX

KW	Atonal; homologue; orthologues; atonal-associated protein; deafness;
KW	hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KM	cellular proliferation; cerebellar granule neuron; gene therapy;
KM	mechanoreceptive cell growth; auditory; osteopathic; cyostatic;
KX	transgenic animal; ss.
OS	Gallus gallus.
XX	
PN	WO200073764-A2.
XX	
PD	07-DEC-2000.
XX	
PF	01-JUN-2000; 2000WO-US15410.
PR	01-JUN-1999; 99US-0137060.
PR	19-JAN-2000; 2000US-0176993.
PA	(BAYU) BAYLOR COLLEGE MEDICINE.
XX	
PI	Zoghbi HY, Bellan H, Birmingham N, Haasan B, Ben-Arie N;
DX	WPI; 2001-032190/04.
DR	P-P5DB; AAB60357.
XX	
PT	Therapeutic use of atonal-associated nucleic acids or amino acids, or
PT	any of its homologs or orthologs, for the treatment of e.g. deafness,
XX	osteoarthritis and abnormal cell proliferation -
XX	
PS	Disclosure; Page -, 142pp; English.
XX	
CC	The invention relates to the use of atonal-associated nucleic acid or
CC	amino acid sequence, or any of its homologues or orthologues as
CC	therapeutic agents for the treatment of deafness, partial hearing loss,
CC	vestibular effects due to damage or loss of inner hair cells,
CC	osteoarthritis and abnormal cell proliferation. The invention also
CC	encompasses methods of screening for compounds which affect the
CC	expression of an atonal-associated nucleic acid sequence in an animal,
CC	and a transgenic animal in which an allele of a native atonal-associated
CC	gene is replaced by a heterologous nucleic acid sequence, thus
CC	inactivating the atonal-associated allele. The nucleic acids or proteins
CC	may be used in a method of treating an animal for hearing impairment,
CC	joint disease, balance disorders, abnormal cell proliferation, or other
CC	disease related to loss of a functional atonal-associated nucleic acid or
CC	protein. They may particularly be used to treat an animal with a
CC	deficiency in cerebellar granule neurons or their precursors, and may
CC	also be used in promoting mechanoreceptive cell growth and generating
CC	hair cells. The present sequence represents an atonal-associated nucleic
CC	acid sequence referred to in the invention.
CC	Note: The present sequence is not shown in the specification, but
CC	was obtained from Genbank.
XX	
SQ	Sequence 790 BP; 91 A; 351 C; 283 G; 65 T; 0 other;
XX	
OY	Query Match 1.6%; Score 23; DB 22; Length 790;
XX	Best Local Similarity 100.0%; Pred.No. 0.65;
XX	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db	732 GAGCGCAACCGCATGCACAACCT 754 374 GAGCGCAACCGCATGCACAACCT 396
RESULT 20	
ID	AAV42938 standard; DNA; 1332 BP.
XX	AAV42938;
XX	25-MAR-2003 (updated)
DT	21-OCT-1998 (first entry)
XX	DNA encoding murine neuroD3 protein, which is a bHLH protein.
DE	
XX	

XX	Basic helix-loop-helix; bHLH; neuroD; neuroectodermal tumour;
KW	classification; medulloblastoma; mouse; ds.
XX	
OS	Mus musculus.
XX	
XX	Key
XX	Location/Qualifiers
XX	101..835
XX	/*tag= a
XX	/product= neuroD3
XX	
XX	US5795723-A.
XX	
XX	18-AUG-1998.
XX	
XX	07-AUG-1997; 97US-0910973.
XX	
XX	06-MAY-1994; 94US-0239238.
XX	02-NOV-1995; 95US-0552142.
XX	30-OCT-1996; 96WO-US17532.
XX	
XX	(HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX	
XX	Olson JM, Tapscott SJ;
XX	
XX	WPI; 1998-466661/40.
XX	P-PSDB; AAW71019.
XX	
XX	Classifying neuroectodermal tumours from expression pattern of
XX	basic-helix-loop-helix genes - especially for identifying
XX	medulla:blastoma and assessing its aggressiveness, specifically
XX	associated with expression of BHLH genes neuroD 1-3
XX	
XX	Example 11; Columns 75-78; 43bp; English.
XX	
XX	The present sequence encodes a protein which is a member of the basic
XX	helix-loop-helix (bHLH) protein family, and is designated neuroD3. The
XX	bHLH genes are a family of genes associated with vertebrate neuronal,
XX	endocrinal and gastrointestinal development. The observed pattern of
XX	neuroD expression distinguishes subclasses of neuroectodermal tumours.
XX	The specification describes a method for the classification of human
XX	neuroectodermal tumours. The method comprises measuring, in a tumour
XX	sample, expression of at least one basic bHLH gene and identifying the
XX	tumour subclass by matching expression to predetermined expression
XX	profiles for known subclasses. For classifying the tumour as a
XX	medulloblastoma, the bHLH gene detected is neuroD and neuroD3.
XX	The method is used to classify neuroectodermal tumours, and to identify
XX	medulloblastoma and for prognosis of this as aggressive.
XX	(Updated on 25-MAR-2003 to correct PR field.)
XX	
XX	Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;
XX	
XX	Query Match 1.6%; Score 23; DB 19; Length 1332;
XX	Best Local Similarity 100.0%; Pred. No. 0.67;
XX	Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX	
XX	732 GAGCGCAACCGCATGCACAACT 754
XX	
XX	404 GAGCGCAACCGCATGCACAACT 426
XX	
XX	RESULT 21
XX	ABS556396
XX	ABS56396 standard; DNA; 1332 BP.
XX	
XX	ABS56396;
XX	
XX	23-JAN-2003 (first entry)
XX	
XX	Mouse bHLH family neuroD3 genomic DNA.
XX	
XX	Mouse; gene; ds; neuroD3; neuroD; basic-helix-loop-helix; bHLH;
XX	differentiation; neuroD; endocrine; gastrointestinal; development;
XX	transgenic; embryo; birth defect; spontaneous abortion; stem cell;
XX	

KW	cancer;neural growth factor; tumor; diagnostic; motor; sensory;
KW	traumatic neural injury; hearing; vision; brain; spinal cord;
KW	malabsorption syndrome; gastrointestinal dysmotility syndrome;
KW	Hirsch Prung's disease; therapeutic.
XX	
OS	Mus musculous.
XX	
FH	Key Location/Qualifiers
FT	CDS 101..835
FT	/tag= a
FT	/product= "NeuroD3"
FT	misc_feature 425..544
FT	/tag= b
FT	/note= "HLH coding domain"
XX	
PN	US6444463-B1.
PD	03-SEP-2002.
PX	
PF	07-FEB-2000; 2000US-0499227.
PR	05-AUG-1998; 98WO-US16417.
PR	08-MAY-1994; 94US-0239238.
PR	06-MAY-1995; 95MO-US05741.
PR	02-NOV-1995; 95US-0552142.
PR	30-OCT-1996; 96WO-US17532.
PR	07-AUG-1997; 97US-0910973.
XX	
PA	(HUTC-) HUTCHINSON CANCER RES CENT FRFD.
XX	
PL	Tapscott SJ;
DR	WP1; 2003-056678/05.
DR	P-PSDB; ABG72005.
XX	
PT	New neurogenic differentiation gene, useful in gene therapy to correct traumatic neural injury that has resulted in loss of motor or sensory neural function and for constructing recombinant cell lines -
PS	Example 11; Column 75-78; 43pp; English.
XX	
CC	The invention discloses an isolated nucleic acid molecule which encodes a functionally active human neuroD3 polypeptide. NeuroD proteins represent a new family within the basic-helix-loop-helix (bHLH) family which are implicated in the regulation of differentiation. NeuroD proteins are particularly involved in neuronal, endocrine and gastrointestinal development. The nucleic acid is useful for constructing recombinant cell lines, transgenic embryos and animals and for quantifying the level of expression of neuroD in a cell. Birth defects and spontaneous abortions may result from expression of an abnormal neuroD protein. The polynucleotide sequences permit the establishment of primary cultures of proliferating embryonic neuronal stem cells under conditions mimicking those that are active in development and cancer. The resultant cell lines find use as sources of novel neural growth factors, in assays for identifying novel neuronal growth factors which can be used for screening anti-cancer drugs capable of driving terminal differentiation in neural tumours, for producing antibodies useful in diagnostic assays and for screening for compounds capable of modulating the activity of neuroD. Transformed host cells, nucleic acids and polypeptides are also useful for treating sites of traumatic neural injury where motor or sensory neural activity has been lost, e.g. hearing or vision loss and brain or spinal cord damage. The host cells find use in the treatment of malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsch Prung's disease). The cell lines also find use in screening for candidate therapeutic agents capable of either substituting for neuroD or correcting the cellular defect caused by a defective neuroD. The sequence presented is the mouse neuroD3 genomic DNA.
SO	Sequence 1332 BP; 268 A; 452 C; 352 G; 260 T; 0 other;

Query Match 1.6%; Score 23; DB 25; Length 1332;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0

QY 732 GAGCGCAACCGCATGCACAACT 754
|||
404 GAGCGCAACCGCATGCACAACT 426
Db

RESULT 22

ID AAT74894 standard; cDNA; 1333 BP

AC AAT74894;

DT 02-OCT-1997 (first entry)

DE Mouse neurogenic differentiation protein (NeuroD3) DNA

Neurogenic differentiation protein; NeuroD; neuroD3 gene;

KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;

XX
WY

. 58

US Mus muscular.
XX

EH	Key	Location/Qualifiers
ET	CNS	101 835

ET
VY

PN W09716548-A1.

PD 09-MAY-1997.

PF 30-OCT-1996; 96WO-US17532.
VY

PR 02-NOV-1995; 95US-0552142.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
PA (HUTN/) WEINTRAUB N

21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855

XX 2007/04

DR P-PSDB; AAW22443.

PT Nucleic acid encoding neurogenic differentiation polypeptide -

PT development

PS Claim 1; Page 72-74; 81pp; English.

CC Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins

CC isolated and sequenced. NeuroD polypeptides are tissue-specific

CC in neuronal, endocrine and gastrointestinal development. They were

cc identify possible PHLH proteins capable of interacting with the
cc discovered by expression cloning and screening assays designed to

protein product of the *Drosophila* *uagnellus* gene; novel neuroblast and neuroD3 genes, related to neuroD1, have been identified.

CC construction of test cell lines as probes, in gene therapy, and for

produce transgenic animals as models of disease.

50 sequence 1333 bp; 268 A; 452 C; 353 G; 260 T; 0 other,

Query Match	1.68;	Score 23;	DB 18;	Length 1333;
Best Local Similarity	100.0%;	Pred. No. 0.67;		

```
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

732 GAGCGCAACCGCATGCACAACCT 754

D_b 404 GAGCGCAACCGCATGCACCAACT 426

```

RESULT 23
AAV27049
ID AAV27049 standard; cDNA; 1385 BP.
XX
XX AAV27049;
AC
XX 17-SEP-1998 (first entry)
XX DT
XX Mouse neurogenin 2 gene.
DE
XX
XX ds; Mouse; neurogenin; expression vector; recombinant protein;
KM antibody; neurogenesis.
XX
XX Mus sp.
OS
XX
XX Key Location/Qualifiers
FH 382..1173
FT /*tag= a
FT /product= "Mouse neurogenin 2"
XX
XX MO9813491-A2.
XX
XX 02-APR-1998.
XX PD
XX
XX 24-SEP-1997; 97WO-US17048.
XX PF
XX
XX 17-SEP-1997; 97US-0932411.
XX PR
XX 27-SEP-1996; 96US-0722570.
XX PR
XX 12-NOV-1996; 96US-0030864.
XX PR
XX 19-DEC-1996; 96US-0772009.
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX PI Anderson DJ, Ma Q, Sommer L;
XX
XX WPI; 1998-230702/20.
XX DR
XX P-PSDB; AAW54946.
XX DR
XX
XX Mouse neurogenins, useful in neurogenesis - and recombinant nucleic
PT acids and proteins derived from rat and Xenopus
XX
XX Claim 5; Fig 7; 106pp; English.
XX PS
XX
XX The Mouse neurogenin 2 is one of several neurogenin proteins discussed
CC in the present invention. The neurogenin nucleic acids can be expressed
CC in a host cell, transformed using an expression vector, to produce
CC recombinant proteins. The proteins and the antibodies raised against
CC the proteins are useful in the study of neurogenesis.
XX
XX SQ Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;

Query Match 1.6%; Score 23; DB 19; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACCT 754
DB 742 GAGCGCAACCGCATGCACAACCT 764

```

```

KM optical nerve damage; auditory nerve damage, neurodegenerative disorder;
KM neuroprotective; nootropic; anticonvulsant; antiParkinsonian; vulnerary;
KM cerebroprotective; immunosuppressant; antiinfectious; de.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
FH 382..1173
FT /*tag= a
FT /product= "Murine neurogenin-2 protein"
XX
XX MO200009676-A2.
XX
XX 24-FEB-2000.
XX PD
XX
XX 13-AUG-1999; 99WO-US18525.
XX PF
XX
XX 14-AUG-1998; 98US-0096630.
XX PR
XX
XX (CALY ) CALIFORNIA INST OF TECHNOLOGY.
XX
XX PI Anderson DJ, Lo L;
XX
XX WPI; 2000-256250/22.
XX DR
XX P-PSDB; AAV70569.
XX DR
XX
XX Inducing non-neuronal cells to differentiate into neurons and for
PT non-neuronal cells to express a neuronal subtype-specific marker,
PT comprising contacting the non-neuronal cells with a vector containing
PT neurogenin nucleic acid -
XX
XX Claim 1; Fig 11; 76pp; English.
XX PS
XX
XX The patent discloses a method for inducing non-neuronal cells (NNC) to
CC differentiate into neurons and for NNCs to express a neuronal subtype
CC -specific marker. Transformed host cells are used as sources of neuronal
CC and other growth factors; in culture for screening compounds that
CC modulate neural differentiation or as sources of recombinantly produced
CC neurogenins and Phox2a proteins for use in transplantation. The cells
CC also have a variety of in vivo uses, e.g. for transplantation at sites of
CC neuronal dysfunction e.g. patients with hearing or vision loss due to
CC optical or auditory nerve damage, brain or spinal cord injuries, and
CC neurodegenerative disorders e.g. Alzheimer's disease. The present
CC sequence encodes murine neurogenin-2 (NGN-2), a transcription factor.
CC NNCs differentiate into neurons through the recombinant expression of a
CC transcription factor that induces a core program of neurogenesis. Forced
CC expression of murine NGN2 can elicit expression of at least some neuronal
CC phenotypic markers even in NNCs.
XX
XX SQ Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;

Query Match 1.6%; Score 23; DB 21; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAACCT 754
DB 742 GAGCGCAACCGCATGCACAACCT 764

```

```

RESULT 24
AAZ51980
ID AAZ51980 standard; DNA; 1385 BP.
XX
XX AAZ51980;
AC
XX
XX 04-JUL-2000 (first entry)
XX DT
XX
XX Murine neurogenin-2 (NGN-2) nucleic acid sequence.
DE
XX
XX Neurogenin-2; NGN-2; non-neuronal cell; NNC; neurogenesis;
KM Phox2a protein; neuronal subtype-specific marker; growth factor;
KM neural differentiation; transplantation; neuronal dysfunction;

```

```

RESULT 25
AAF27269
ID AAF27269 standard; cDNA; 1385 BP.
XX
XX AAF27269;
AC
XX
XX 24-APR-2001 (first entry)
XX DT
XX
XX Mouse neurogenin 2 (ngn2) cDNA, SEQ ID NO:30.
DE
XX
XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
KM hearing impairment; vestibular effect; balance disorder; osteoarthritis;
KM cellular proliferation; cerebellar granule neuron; gene therapy;

```


KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.
 XX Mus musculus.
 XX
 XX WO200073764-A2.
 XX
 XX 07-DEC-2000.
 XX
 XX 01-JUN-2000; 2000WO-US15410.
 XX
 XX 01-JUN-1999; 99US-0137060.
 XX
 XX 19-JAN-2000; 2000US-0176993.
 XX
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 PA
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 PI
 XX WPI; 2001-032190/04.
 XX
 XX P-PSDB; AAB60362.
 DR
 XX
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness, or
 PT osteoarthritis and abnormal cell proliferation -
 PT
 XX Disclosure; Page -; 142pp; English.
 PS
 XX The invention relates to the use of atonal-associated nucleic acid or
 XX amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 CC
 XX
 XX Sequence 1385 BP; 242 A; 467 C; 432 G; 244 T; 0 other;
 SQ
 Query Match 1.6%; Score 23; DB 22; Length 1385;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGGCAACCGCATGCACAACCT 754
 |||||
 Db 742 GAGGCAACCGCATGCACAACCT 764

RESULT 26
 AAF27255
 ID AAF27255 standard; CDNA; 1412 BP.
 XX
 XX AAF27255;
 AC
 XX 24-APR-2001 (first entry)
 DT
 XX Mouse atonal homologue 4 (ATOH4, Math4A) cDNA, SEQ ID NO:6.
 DE
 XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;

KW transgenic animal; ss.
 XX
 XX Mus musculus.
 XX
 XX WO200073764-A2.
 XX
 XX 07-DEC-2000.
 XX
 XX 01-JUN-2000; 2000WO-US15410.
 XX
 XX 01-JUN-1999; 99US-0137060.
 XX
 XX 19-JAN-2000; 2000US-0176993.
 XX
 XX (BAYU) BAYLOR COLLEGE MEDICINE.
 PA
 XX Zoghbi HY, Bellen H, Birmingham N, Hassan B, Ben-Arie N;
 PI
 XX WPI; 2001-032190/04.
 XX
 XX P-PSDB; AAB60351.
 DR
 XX
 XX Therapeutic use of atonal-associated nucleic acids or amino acids, or
 PT any of its homologs or orthologs, for the treatment of e.g. deafness, or
 PT osteoarthritis and abnormal cell proliferation -
 PT
 XX Disclosure; Page -; 142pp; English.
 PS
 XX The invention relates to the use of atonal-associated nucleic acid or
 XX amino acid sequence, or any of its homologues or orthologues as
 CC therapeutic agents for the treatment of deafness, partial hearing loss,
 CC vestibular effects due to damage or loss of inner hair cells,
 CC osteoarthritis and abnormal cell proliferation. The invention also
 CC encompasses methods of screening for compounds which affect the
 CC expression of an atonal-associated nucleic acid sequence in an animal,
 CC and a transgenic animal in which an allele of a native atonal-associated
 CC gene is replaced by a heterologous nucleic acid sequence, thus
 CC inactivating the atonal-associated allele. The nucleic acids or proteins
 CC may be used in a method of treating an animal for hearing impairment,
 CC joint disease, balance disorders, abnormal cell proliferation, or other
 CC disease related to loss of a functional atonal-associated nucleic acid or
 CC protein. They may particularly be used to treat an animal with a
 CC deficiency in cerebellar granule neurons or their precursors, and may
 CC also be used in promoting mechanoreceptive cell growth and generating
 CC hair cells. The present sequence represents an atonal-associated nucleic
 CC acid sequence referred to in the invention.
 CC Note: The present sequence is not shown in the specification, but
 CC was obtained from GenBank.
 CC
 XX
 XX Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;
 SQ
 Query Match 1.6%; Score 23; DB 22; Length 1412;
 Best Local Similarity 100.0%; Pred. No. 0.67;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGGCAACCGCATGCACAACCT 754
 |||||
 Db 445 GAGGCAACCGCATGCACAACCT 467

RESULT 27
 AAF27273
 ID AAF27273 standard; CDNA; 1412 BP.
 XX
 XX AAF27273;
 AC
 XX 24-APR-2001 (first entry)
 DT
 XX Mouse atonal homologue 4A (Math4A) cDNA, SEQ ID NO:37.
 DE
 XX Atonal; homologue; orthologue; atonal-associated protein; deafness;
 KW hearing impairment; vestibular effect; balance disorder; osteoarthritis;
 KW cellular proliferation; cerebellar granule neuron; gene therapy;
 KW mechanoreceptive cell growth; auditory; osteopathic; cytostatic;
 KW transgenic animal; ss.

```

XX OS Mus musculus.
XX PN WO200073764-A2.
XX PD 07-DEC-2000.
XX PF 01-JUN-2000; 2000WO-US15410.
XX PR 01-JUN-1999; 99US-0137060.
XX PR 19-JAN-2000; 2000US-0176993.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PI Zoghbi HY, Bellen H, Birmingham N, Haasan B, Ben-Arie N;
XX WPI: 2001-032190/04.
XX DR P-PSDB; AAB60365.
XX PT Therapeutic use of atonal-associated nucleic acids or amino acids, or
XX PT any of its homologs or orthologs, for the treatment of e.g. deafness,
XX PT osteoarthritis and abnormal cell proliferation -
XX PS Disclosure; Page -: 142pp; English.
XX CC The invention relates to the use of atonal-associated nucleic acid or
XX CC amino acid sequence, or any of its homologues or orthologues as
XX CC therapeutic agents for the treatment of deafness, partial hearing loss,
XX CC vestibular effects due to damage or loss of inner hair cells,
XX CC osteoarthritis and abnormal cell proliferation. The invention also
XX CC encompasses methods of screening for compounds which affect the
XX CC expression of an atonal-associated nucleic acid sequence in an animal,
XX CC and a transgenic animal in which an allele of a native atonal-associated
XX CC gene is replaced by a heterologous nucleic acid sequence, thus
XX CC inactivating the atonal-associated allele. The nucleic acids or proteins
XX CC may be used in a method of treating an animal for hearing impairment,
XX CC joint disease, balance disorders, abnormal cell proliferation, or other
XX CC disease related to loss of a functional atonal-associated nucleic acid or
XX CC protein. They may particularly be used to treat an animal with a
XX CC deficiency in cerebellar granule neurons or their precursors, and may
XX CC also be used in promoting mechanoreceptive cell growth and generating
XX CC hair cells. The present sequence represents an atonal-associated nucleic
XX CC acid sequence referred to in the invention.
XX CC Note: The present sequence is not shown in the specification, but
XX CC was obtained from GenBank.
XX SQ Sequence 1412 BP; 265 A; 447 C; 435 G; 265 T; 0 other;

Query Match 1.6%; Score 23; DB 22; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACT 754
DB 445 GAGCGCAACCGCATGCACACT 467

RESULT 28
AADD47278
XX ID AAD47278 standard; DNA; 21 BP.
XX AC AAD47278;
XX DT 24-FEB-2003 (first entry)
XX DE Human RT-PCR upstream primer for neurogenin-3 DNA isolation.
XX KW Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
XX KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX OS Homo sapiens.
XX PA (NOVO ) NOVO NORDISK AS.
XX PI Serup P, Heimberg H, Gradwohl G;
XX WPI: 20020274946-A2.

```

```

XX XX 26-SEP-2002.
XX PD 26-FEB-2002; 2002WO-DK00130.
XX PF 26-FEB-2001; 2001US-271474P.
XX PR (NOVO ) NOVO NORDISK AS.
XX PA (NOVO ) NOVO NORDISK AS.
XX PI Serup P, Heimberg H, Gradwohl G;
XX WPI: 2003-018804/01.
XX DR WPI: 2003-018804/01.
XX PT Generating insulin-secreting cells from precursor stem cells or adult
XX PT pancreatic exocrine cells, for generating glucose sensitive insulin
XX PT secreting beta cells for transplantation, comprises using neurogenin3
XX PT or NeuroD/beta2 -
XX PS Example 4; Page 28; 66pp; English.
XX CC The invention relates to a method for generating insulin-secreting cells
XX CC from precursor stem cells or adult pancreatic exocrine cells. The method
XX CC comprises exposing the precursor cells or exocrine cells to a nucleic
XX CC acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an
XX CC activator of ngn3 or NeuroD/beta2 gene expression, under conditions
XX CC effective to generate the insulin-generating cells from the precursor or
XX CC exocrine cells. The invention is useful in generating insulin-secreting
XX CC cells from precursor stem cells or adult pancreatic exocrine cells is
XX CC suitable for generating glucose sensitive insulin secreting beta cells
XX CC secreting cells in a patient. The method is also useful for preventing
XX CC premature differentiation of precursor stem cells into insulin-secreting
XX CC beta cells and for identifying compounds that prevent or activate beta
XX CC cell differentiation. The present sequence is human RT-PCR primer for
XX CC isolation of neurogenin-3 DNA.
XX SQ Sequence 21 BP; 2 A; 7 C; 6 G; 6 T; 0 other;

Query Match 1.4%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 460 TGGCGCCTCATCCCTTGATG 480
DB 1 TGGCGCCTCATCCCTTGATG 21

RESULT 29
AADD47326
XX ID AAD47326 standard; DNA; 21 BP.
XX AC AAD47326;
XX DT 24-FEB-2003 (first entry)
XX DE Human RT-PCR forward primer for mouse ngn3 DNA isolation.
XX KW Human; insulin-secreting cell; neurogenin 3; ngn3; precursor stem cell;
XX KW pancreatic exocrine cell; transplantation; RT-PCR; primer; ss.
XX OS Homo sapiens.
XX PA (NOVO ) NOVO NORDISK AS.
XX PI Serup P, Heimberg H, Gradwohl G;
XX WPI: 20020274946-A2.
XX PD 26-SEP-2002.
XX PF 26-FEB-2002; 2002WO-DK00130.
XX PR 26-FEB-2001; 2001US-271474P.
XX PA (NOVO ) NOVO NORDISK AS.
XX PI Serup P, Heimberg H, Gradwohl G;
XX WPI: 20020274946-A2.

```

XX WPI; 2003-018804/01.
XX
XX Generating insulin-secreting cells from precursor stem cells or adult
PT pancreatic exocrine cells, for generating glucose sensitive insulin
PT secreting beta cells for transplantation, comprises using neurogenin3
PT or NeuroD/beta2 -
XX
XX Example 5B; Page 37; 66pp; English.
XX
XX The invention relates to a method for generating insulin-secreting cells
CC from precursor stem cells or adult pancreatic exocrine cells. The method
CC comprises exposing the precursor cells or exocrine cells to: a nucleic
CC acid molecule encoding neurogenin 3 (ngn3) or NeuroD/beta2; or an
CC activator of ngn3 or NeuroD/beta2 gene expression, under conditions
CC effective to generate the insulin-generating cells from the precursor or
CC exocrine cells. The invention is useful in generating insulin-secreting
CC cells from precursor stem cells or adult pancreatic exocrine cells is
CC useful for generating glucose sensitive insulin secreting beta cells
CC suitable for transplantation, and for in situ development of insulin-
CC secreting cells in a patient. The method is also useful for preventing
CC premature differentiation of precursor stem cells into insulin-secreting
CC beta cells and for identifying compounds that prevent or activate beta
CC cell differentiation. The present sequence is human RT-PCR primer for
CC isolation of mouse ngn3 DNA.
XX
XX Sequence 21 BP; 7 A; 7 C; 4 G; 3 T; 0 other;
SQ
Query Match 1.4%; Score 21; DB 25; Length 21;
Best Local Similarity 100.0%; Pred. No. 7.7;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 798 CCGGATGACCGCAACTTACA 818
Db 1 CCGGATGACCGCAACTTACA 21
RESULT 30
AAD46889
ID AAD46889 standard; DNA; 714 BP.
XX
XX AAD46889;
XX
XX 27-JAN-2003 (first entry)
XX
XX Human neurogenin 1 (Ngn1) gene #2.
XX
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
KW type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
KW islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 1..714
FT /*tag= a
FT /product= "Human Ngn1 protein"
XX
XX WO200274045-A2.
XX
XX 26-SEP-2002.
XX
XX 20-MAR-2002; 2002WO-US11166.
XX
XX 20-MAR-2001; 2001US-0817360.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX German MS, Lin J;
XX
XX WPI; 2002-755953/82.
XX
XX P-PSDB; AAE29280.
XX

PT Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
XX Disclosure; Page 94; 108pp; English.
XX
XX The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 1
CC (Ngn1) gene.
XX
XX Sequence 714 BP; 118 A; 287 C; 207 G; 102 T; 0 other;
SQ
Query Match 1.4%; Score 21; DB 24; Length 714;
Best Local Similarity 100.0%; Pred. No. 6.6;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 732 GAGCGCAACCGCATGCACAC 752
Db 301 GAGCGCAACCGCATGCACAC 321
RESULT 31
AAT74891
ID AAT74891 standard; DNA; 1268 BP.
XX
XX AAT74891;
XX
XX 02-OCT-1997 (first entry)
XX
XX Human neurogenic differentiation protein (NeuroD) DNA clone 20A1.
XX
XX Neurogenic differentiation protein; NeuroD; neuroD3 gene;
KW transcriptional activator; neuron; pancreas; gastrointestinal;
KW knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;
KW ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 55..768
FT /*tag= a
FT /product= "Human NeuroD protein"
XX
XX WO9716548-A1.
XX
XX 09-MAY-1997.
XX
XX 30-OCT-1996; 96WO-US17532.
XX
XX 02-NOV-1995; 95US-0552142.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX (WEIN/) WEINTRAUB N.
XX
XX Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;
XX
XX WPI; 1997-272117/24.
XX
XX P-PSDB; AAW22440.
XX
XX Nucleic acid encoding neurogenic differentiation polypeptide -
PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
PT development
XX
XX Claim 1; Page 64-65; 81pp; English.
XX
XX Neurogenic differentiation (NeuroD) genes (AAT74887-94) and proteins
CC (AAW22436-43) from human, mouse and frog have been identified.

CC isolated and sequence. Neurod polypeptides are tissue-specific
CC basic-helix-loop-helix (bHLH) transcriptional activators involved
CC in neuronal, endocrine and gastrointestinal development. They were
CC discovered by expression cloning and screening assays designed to
CC identify possible bHLH proteins capable of interacting with the
CC protein product of the Drosophila daughterless gene. Novel neurod2
CC and neurod3 genes, related to neurod1, have been identified.
CC Neurod nucleic acids can be used to produce Neurod polypeptides,
CC construction of test cell lines, as probes, in gene therapy, and to
CC produce transgenic animals as models of disease.

XX SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;
XX
XX Query Match 1.4%; Score 21; DB 18; Length 1268;
XX Best Local Similarity 100.0%; Pred. No. 6.5;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752
DB 355 GAGCGCAACCGCATGCACAC 375
|||||
|||||

RESULT 32
AAV42932
ID AAV42932 standard; DNA; 1268 BP.
XX
XX AAV42932;
XX
XX 25-MAR-2003 (updated)
XX 21-OCT-1998 (first entry)
XX
XX DNA encoding human neurod3 protein, which is a bHLH protein.
XX
XX Basic helix-loop-helix; bHLH; neurod; neuroectodermal tumour;
XX classification; medulloblastoma; human; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 55..768
XX /*tag= a
XX /product= neurod3
XX
XX US5795723-A.
XX
XX 18-AUG-1998.
XX
XX 07-AUG-1997; 97US-0910973.
XX
XX 06-MAY-1994; 94US-0239238.
XX 02-NOV-1995; 95US-0552142.
XX 30-OCT-1996; 96WO-US17532.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Olson JM, Tapscott SJ;
XX
XX WPI: 1998-466661/40.
XX P-PSDB; AAW71016.
XX
XX Classifying neuroectodermal tumours from expression pattern of
XX basic-helix-loop-helix genes - especially for identifying
XX medulla:blastoma and assessing its aggressiveness, specifically
XX associated with expression of bHLH genes neurod 1-3
XX
XX Example 11; Columns 61-64; 43pp; English.
XX
XX The present sequence encodes a protein which is a member of the basic
XX helix-loop-helix (bHLH) protein family, and is designated neurod3. The
XX bHLH genes are a family of genes associated with vertebrate neuronal,
XX endocrinal and gastrointestinal development. The observed pattern of
XX neurod expression distinguishes subclasses of neuroectodermal tumours.
XX The specification describes a method for the classification of human

CC neuroectodermal tumours. The method comprises measuring, in a tumour
CC sample, expression of at least one basic bHLH gene and identifying the
CC tumour subclasses by matching expression to predetermined expression
CC profiles for known subclasses. For classifying the tumour as a
CC medulloblastoma, the bHLH gene detected is neurod1 and neurod3.
CC The method is used to classify neuroectodermal tumours, and to identify
CC medulloblastoma and for prognosis of this as aggressive.
CC (updated on 25-MAR-2003 to correct PR field.)
XX
XX SQ Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;
XX
XX Query Match 1.4%; Score 21; DB 19; Length 1268;
XX Best Local Similarity 100.0%; Pred. No. 6.5;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACAC 752
DB 355 GAGCGCAACCGCATGCACAC 375
|||||
|||||

RESULT 33
ABS56390
ID ABS56390 standard; DNA; 1268 BP.
XX
XX ABS56390;
XX
XX 23-JAN-2003 (first entry)
XX
XX Human bHLH family neurod3 genomic DNA, clone 20A1.
XX
XX Human; gene; ds; neurod3; neurod; basic-helix-loop-helix; bHLH;
XX differentiation; neurone; endocrine; gastrointestinal; development;
XX transgenic; embryo; birth defect; spontaneous abortion; stem cell;
XX cancer; neural growth factor; tumour; diagnostic; motor; sensory;
XX traumatic neural injury; hearing; vision; brain; spinal cord;
XX malabsorption syndrome; gastrointestinal dysmotility syndrome;
XX Hirschprung's disease; therapeutic.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 55..768
XX /*tag= a
XX /product= "Neurod3"
XX
XX misc_feature 376..495
XX /*tag= b
XX /note= "HLH coding domain"
XX
XX US6444463-B1.
XX
XX 03-SEP-2002.
XX
XX 07-FEB-2000; 2000US-0499227.
XX
XX 05-AUG-1998; 98WO-US16417.
XX 06-MAY-1994; 94US-0239238.
XX 08-MAY-1995; 95WO-US05741.
XX 02-NOV-1995; 95US-0552142.
XX 30-OCT-1996; 96WO-US17532.
XX 07-AUG-1997; 97US-0910973.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
XX Tapscott SJ;
XX
XX WPI: 2003-056678/05.
XX P-PSDB; ABG72002.
XX
XX New neurogenic differentiation gene, useful in gene therapy to correct
XX traumatic neural injury that has resulted in loss of motor or sensory
XX neural function and for constructing recombinant cell lines -
XX
XX Claim 1; Column 61-64; 43pp; English.

XX The invention discloses an isolated nucleic acid molecule which encodes a
 CC functionally active human neurod3 polypeptide. Neurod proteins represent
 CC a new family within the basic-helix-loop-helix (bHLH) family which are
 CC implicated in the regulation of differentiation. Neurod proteins are
 CC particularly involved in neuronal, endocrine and gastrointestinal
 CC development. The nucleic acid is useful for constructing recombinant cell
 CC lines, transgenic embryos and animals and for quantifying the level of
 CC expression of neurod in a cell. Birth defects and spontaneous abortions
 CC may result from expression of an abnormal neurod protein. The
 CC polynucleotide sequences permit the establishment of primary cultures of
 CC proliferating embryonic neuronal stem cells under conditions mimicking
 CC those that are active in development and cancer. The resultant cell lines
 CC find use as sources of novel neural growth factors, in assays for
 CC identifying novel neuronal growth factors which can be used for screening
 CC anti-cancer drugs capable of driving terminal differentiation in neural
 CC tumours, for producing antibodies useful in diagnostic assays and for
 CC screening for compounds capable of modulating the activity of neurod.
 CC Transformed host cells, nucleic acids and polypeptides are also useful
 CC for treating sites of traumatic neural injury where motor or sensory
 CC neural activity has been lost, e.g. hearing or vision loss and brain or
 CC spinal cord damage. The host cells find use in the treatment of
 CC malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsch
 CC Prung's Disease). The cell lines also find use in screening for candidate
 CC therapeutic agents capable of either substituting for neurod or
 CC correcting the cellular defect caused by a defective neurod. The sequence
 CC presented is the human neurod3 genomic DNA, clone 20A1.

SO Sequence 1268 BP; 245 A; 455 C; 344 G; 224 T; 0 other;

Query Match 1.4%; Score 21; DB 25; Length 1268;
 Best Local Similarity 100.0%; Pred. No. 6.5;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGGCAACCGCATGCACAC 752
 Db 355 GAGGCAACCGCATGCACAC 375

RESULT 34

AAAT74890 standard; DNA; 1535 BP.

AC AAT74890;

DT 02-OCT-1997 (first entry)

DE Human neurogenic differentiation protein (Neurod2) DNA clone 14B1.

XX Neurogenic differentiation protein; Neurod; neurod2 gene;

KM transcriptional activator; neuron; pancreas; gastrointestinal;

KM knock-out mouse; transgenic animal; cancer; diabetes; gene therapy;

ss.

XX Homo sapiens.

OS Location/Qualifiers

FT Key 55..1200

FT CDS /*tag= a

XX MO9716548-A1.

XX 09-MAY-1997.

XX 30-OCT-1996; 96WO-US17532.

XX 02-NOV-1995; 95US-0552142.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.
 (WEIN/) WEINTRAUB N.

PI Hollenberg SM, Lee JE, Tapscott SJ, Weintraub HM;

DR WPI; 1997-272117/24.
 DR P-PSDB; AAM22439.
 XX Nucleic acid encoding neurogenic differentiation polypeptide -
 PT useful e.g. in regulating neuronal, endocrine and gastrointestinal
 PT development
 PS Claim 1; Page 61-62; 81pp; English.
 XX Neurogenic differentiation (Neurod) genes (AAT74887-94) and proteins
 CC (AAM22436-43) from human, mouse and frog have been identified;
 CC isolated and sequenced. Neurod polypeptides are tissue-specific
 CC basic-helix-loop-helix (bHLH) transcriptional activators involved
 CC in neuronal, endocrine and gastrointestinal development. They were
 CC discovered by expression cloning and screening assays designed to
 CC identify possible bHLH proteins capable of interacting with the
 CC protein product of the Drosophila daughterless gene. Novel neurod2
 CC and neurod3 genes, related to neurod1, have been identified.
 CC Neurod nucleic acids can be used to produce Neurod polypeptides,
 CC construction of test cell lines, as probes, in gene therapy, and to
 CC produce transgenic animals as models of disease.

SO Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;

Query Match 1.4%; Score 21; DB 18; Length 1535;
 Best Local Similarity 100.0%; Pred. No. 6.4;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 729 CGGAGCGCAACCGCATGCAC 749
 Db 439 CGGAGCGCAACCGCATGCAC 459

RESULT 35

AAV42931 standard; DNA; 1535 BP.

AC AAV42931;

DT 25-MAR-2003 (updated)

DT 21-OCT-1998 (first entry)

DE DNA encoding human neurod2 protein, which is a bHLH protein.

XX Basic helix-loop-helix; bHLH; neurod; neuroectodermal tumour;

KM classification; medulloblastoma; human; ds.

XX Homo sapiens.

OS Location/Qualifiers

FT Key 55..1194

FT CDS /*tag= a

FT /*product= neurod2

XX US5795723-A.

XX 18-AUG-1998.

XX 07-AUG-1997; 97US-0910973.

XX 06-MAY-1994; 94US-0239238.

XX 02-NOV-1995; 95US-0552142.

XX 30-OCT-1996; 96WO-US17532.

PA (HUTC-) HUTCHINSON CANCER RES CENT FRED.

PI Olson JM, Tapscott SJ;

DR WPI; 1998-466661/40.

DR P-PSDB; AAM71015.

PT Classifying neuroectodermal tumours from expression pattern of
 basic-helix-loop-helix genes - especially for identifying

PT medulla:blastoma and assessing its aggressiveness, specifically
PT associated with expression of BHLH genes neurod 1-3
XX
PS Example 11; Columns 57-60; 43pp; English.
XX
CC The present sequence encodes a protein which is a member of the basic
CC helix-loop-helix (bHLH) protein family, and is designated neurod2. The
CC bHLH genes are a family of genes associated with vertebrate neuronal,
CC endocrinal and gastrointestinal development. The observed pattern of
CC neurod expression distinguishes subclasses of neuroectodermal tumours.
CC The specification describes a method for the classification of human
CC neuroectodermal tumours. The method comprises measuring, in a tumour
CC sample, expression of at least one basic bHLH gene and identifying the
CC tumour subclasses by matching expression to predetermined expression
CC profiles for known subclasses. For classifying the tumour as a
CC medulloblastoma, the bHLH gene detected is neurod1 and neurod3.
CC The method is used to classify neuroectodermal tumours, and to identify
CC medulloblastoma and for prognosis of this as aggressive.
CC (Updated on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;
Query Match 1.4%; Score 21; DB 19; Length 1535;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 729 CGGAGCGCAACCGCATGCAC 749
Db 439 CGGAGCGCAACCGCATGCAC 459
RESULT 36
ABS56389
ID ABS56389 standard; DNA; 1535 BP.
XX
AC ABS56389;
XX
DT 23-JAN-2003 (first entry)
XX
DE Human bHLH family neurod2 genomic DNA, clone 14B1.
XX
KW Human; gene; ds; neurod3; neurod; basic-helix-loop-helix; bHLH;
KW differentiation; neurone; endocrine; gastrointestinal; development;
KW transgenic; embryo; birth defect; spontaneous abortion; stem cell;
KW cancer; neural growth factor; tumour; diagnostic; motor; sensory;
KW traumatic neural injury; hearing; vision; brain; spinal cord;
KW malabsorption syndrome; gastrointestinal dysmotility syndrome;
KW Hirsh Prung's disease; therapeutic; neurod2.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 55..1194
FT /*tag= a
FT /product= "Neurod2"
FT misc_feature 463..582
FT /*tag= b
FT /note= "HLH coding domain"
XX
PN US6444463-B1.
XX
PD 03-SEP-2002.
XX
PF 07-FEB-2000; 2000US-0499227.
XX
XX 05-AUG-1998; 98WO-US16417.
PR 06-MAY-1994; 94US-0239238.
PR 08-MAY-1995; 95US-US05741.
PR 02-NOV-1995; 95US-0552142.
PR 30-OCT-1996; 96WO-US17532.
PR 07-AUG-1997; 97US-0910973.
XX
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.

XX
PI Tapscott SJ;
XX
DR WPI; 2003-056678/05.
DR P-PSDB; ABG72001.
XX
XX
PT New neurogenic differentiation gene, useful in gene therapy to correct
PT traumatic neural injury that has resulted in loss of motor or sensory
PT neural function and for constructing recombinant cell lines
XX
PS Example 11; Column 57-60; 43pp; English.
XX
CC The invention discloses an isolated nucleic acid molecule which encodes a
CC functionally active human neurod3 polypeptide. Neurod proteins represent
CC a new family within the basic-helix-loop-helix (bHLH) family which are
CC implicated in the regulation of differentiation. Neurod proteins are
CC particularly involved in neuronal, endocrine and gastrointestinal
CC development. The nucleic acid is useful for constructing recombinant cell
CC lines, transgenic embryos and animals and for quantifying the level of
CC expression of neurod in a cell. Birth defects and spontaneous abortions
CC may result from expression of an abnormal neurod protein. The
CC polynucleotide sequences permit the establishment of primary cultures of
CC proliferating embryonic neuronal stem cells under conditions mimicking
CC those that are active in development and cancer. The resultant cell lines
CC find use as sources of novel neural growth factors, in assays for
CC identifying novel neuronal growth factors which can be used for screening
CC anti-cancer drugs capable of driving terminal differentiation in neural
CC tumours, for producing antibodies useful in diagnostic assays and for
CC screening for compounds capable of modulating the activity of neurod.
CC Transformed host cells, nucleic acids and polypeptides are also useful
CC for treating sites of traumatic neural injury where motor or sensory
CC neural activity has been lost, e.g. hearing or vision loss and brain or
CC spinal cord damage. The host cells find use in the treatment of
CC malabsorption syndromes or gastrointestinal dysmotility syndromes (Hirsh
CC Prung's Disease). The cell lines also find use in screening for candidate
CC therapeutic agents capable of either substituting for neurod or
CC correcting the cellular defect caused by a defective neurod. The sequence
CC presented is the human neurod2 genomic DNA, clone 14B1.
XX
SQ Sequence 1535 BP; 250 A; 559 C; 476 G; 244 T; 6 other;
Query Match 1.4%; Score 21; DB 25; Length 1535;
Best Local Similarity 100.0%; Pred. No. 6.4;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 729 CGGAGCGCAACCGCATGCAC 749
Db 439 CGGAGCGCAACCGCATGCAC 459
RESULT 37
AAA62681
ID AAA62681 standard; DNA; 1550 BP.
XX
AC AAA62681;
XX
DT 29-NOV-2000 (first entry)
XX
DE Human Neurod2 gene.
XX
KW Human; Neurod2; neurogenic basic helix-loop-helix protein;
KW epidermal cell transdifferentiation; gene therapy; cerebroprotective;
KW neuroprotective; brain injury; spinal cord injury; stroke;
KW neurodegenerative disease; Parkinson's disease; Huntington's disease;
KW Alzheimer's disease; neuronal cell generation; ds.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 58..1206
FT /*tag= a
FT /product= "Neurod2 protein"
XX

```
PN US6087168-A.
XX
XX 11-JUL-2000.
XX
XX 20-JAN-1999; 99US-0234332.
XX
XX 20-JAN-1999; 99US-0234332.
XX
XX (CEDA-) CEDARS SINAI MEDICAL CENT.
XX
XX Levesque MF, Neuman T;
XX
XX WPI; 2000-498200/44.
XX
XX P-PSDB; AAB14347.
XX
XX Converting epidermal cells into neurons, useful for isolating nerve
PT growth factors or for gene therapy, comprises differentiating cells
PT and transfecting with vectors with a cDNA coding neurogenic
PT transcription factors -
XX
XX Example 2; Column 19-22; 27pp; English.
XX
XX The present sequence is the human neurogenic basic helix-loop-helix
CC protein (NeuroD2) gene from Genbank. It was used to provide sequence
CC information for the cloning of NeuroD2 cDNA, which was used to
CC transfect cultured epidermal cells. This was part of a novel method
CC for transdifferentiating an epidermal basal cell into a cell having the
CC morphological, physiological and/or immunological features of a viable
CC neuronal cell. The method is useful for screening new drugs for
CC treating a nervous system disorder, or for isolating a novel nerve
CC growth factor. The transdifferentiated cell is useful in both cell and
CC cell or gene therapy aimed at alleviating various neurological disorders. The
CC transplantation or grafting of the newly created neuronal cells as
CC treatment for brain or spinal cord injury, stroke and neurodegenerative
CC diseases (e.g. Parkinson's disease, Huntington's disease or Alzheimer's
CC disease).
XX
XX Sequence 1550 BP; 250 A; 569 C; 486 G; 243 T; 2 other;
SQ
XX
XX Query Match 1.4%; Score 21; DB 21; Length 1550;
XX Best Local Similarity 100.0%; Pred. No. 6.4;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 729 CGGAGCGCAACCGCATGCAC 749
DB 442 CGGAGCGCAACCGCATGCAC 462
RESULT 38
AAD46888 standard; DNA; 1665 BP.
XX
XX AAD46888;
XX
XX 27-JAN-2003 (first entry)
XX
XX Human neurogenin 1 (Ngn1) gene #1.
XX
XX Human; transcription factor; neuroendocrine basic helix-loop-helix; bHLH;
XX type 2 diabetes mellitus; autoimmune destruction; type 1 diabetes;
XX islet cell; cell therapy; neurogenin 1; Ngn1; gene; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 260..973
XX /tag= a
XX /product= "Human Ngn1 protein"
XX
XX MO200274045-A2.
XX
XX 26-SEP-2002.
XX
```

```
XX
XX 20-MAR-2002; 2002WO-US11166.
XX
XX 20-MAR-2001; 2001US-0817360.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX German MS, Lin J;
XX
XX WPI; 2002-759853/82.
XX
XX P-PSDB; AAE29279.
XX
XX Producing a mammalian islet cell for treating diabetes mellitus
PT comprises introducing into a mammalian cell a nucleic acid molecule
PT encoding neuroendocrine basic helix-loop-helix transcription factor -
XX
XX Disclosure; Page 93; 108pp; English.
XX
XX The invention relates to a method for producing a mammalian islet cell.
CC The method comprising introducing into a mammalian cell a nucleic acid
CC molecule encoding an islet transcription factor for expression of the
CC islet transcription factor in the cell and for production of islet cell
CC phenotype in the cell. The islet transcription factor is a neuroendocrine
CC basic helix-loop-helix (bHLH) transcription factor. The method is useful
CC for treating type 2 diabetes mellitus and for replacing beta cells lost
CC to autoimmune destruction in individuals with type 1 diabetes. The method
CC is useful in cell therapy. The present sequence is human neurogenin 1
CC (Ngn1) gene.
XX
XX Sequence 1665 BP; 344 A; 561 C; 443 G; 317 T; 0 other;
SQ
XX
XX Query Match 1.4%; Score 21; DB 24; Length 1665;
XX Best Local Similarity 100.0%; Pred. No. 6.4;
XX Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 732 GAGCGCAACCGCATGCACAC 752
DB 560 GAGCGCAACCGCATGCACAC 580
RESULT 39
AAL04043/c
XX AAL04043 standard; DNA; 2776 BP.
XX
XX AAL04043;
XX
XX 21-NOV-2001 (first entry)
XX
XX Human reproductive system related antigen DNA SEQ ID NO: 6731.
XX
XX Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
XX Homo sapiens.
XX
XX MO200155320-A2.
XX
XX 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01339.
XX
XX 31-JAN-2000; 2000US-0179065.
XX
XX 04-FEB-2000; 2000US-0180628.
XX
XX 24-FEB-2000; 2000US-0184664.
XX
XX 02-MAR-2000; 2000US-0186350.
XX
XX 16-MAR-2000; 2000US-0189874.
XX
XX 17-MAR-2000; 2000US-0190076.
XX
XX 18-APR-2000; 2000US-0198123.
XX
XX 19-MAY-2000; 2000US-0205515.
XX
XX 07-JUN-2000; 2000US-0209467.
XX
XX 28-JUN-2000; 2000US-0214886.
XX
XX 30-JUN-2000; 2000US-0215135.
XX
XX 07-JUL-2000; 2000US-0216647.
XX
```

[illegible]

PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.
PR	08-NOV-2000;	2000US-0246532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249299.
PR	01-DEC-2000;	2000US-0250300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0251719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-465576/50.	
DR		
XX		
PT	Isolated nucleic acid molecule encoding a reproductive system antigen	
XX	is used in preventing, treating or ameliorating a medical condition -	
PT		
XX		
PS	Disclosure; SEQ ID NO 6731; 1297pp + Sequence Listing; English.	
XX		
CC	The present invention provides the protein and coding sequences of a	
CC	number of human reproductive system related antigens. These can be used	
CC	in the prevention and treatment of reproductive system disorders,	
CC	including cancer. The present sequence is a genomic sequence encoding a	
CC	protein of the invention.	
XX		
SQ	Sequence 2776 BP; 631 A; 851 C; 749 G; 545 T; 0 other:	
Query Match	1.4%; Score 21; DB 22; Length 2776;	
Best Local Similarity	100.0%; Pred. No. 6.3;	


```
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1112 AACAGCCCTGGCGGTGGGC 1132
Db 938 AACAGCCCTGGCGGTGGGC 918

RESULT 40
AAL04045/C
ID AAL04045 standard; DNA; 2776 BP.
XX
AC AAL04045;
XX
DT 21-NOV-2001 (first entry)
XX
DE Human reproductive system related antigen DNA SEQ ID NO: 6733.
XX
KW Human; reproductive system related antigen; reproductive system disorder;
XX cancer; gene therapy; ds.
XX
OS Homo sapiens.
XX
PN WO200155320-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01339.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225276.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 08-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
```

```
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232060.
PR 08-SEP-2000; 2000US-0232061.
PR 12-SEP-2000; 2000US-0232081.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
```

PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249266.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249269.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250191.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251899.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCT INC.
 XX
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-465570/50.
 XX
 PS Disclosure; SEQ ID NO 6733; 1297pp + Sequence Listing; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of human reproductive system related antigens. These can be used
 CC in the prevention and treatment of reproductive system disorders,
 CC including cancer. The present sequence is a genomic sequence encoding a
 CC protein of the invention.
 XX
 XX Sequence 2776 BP; 631 A; 853 C; 747 G; 545 T; 0 other;

Query Match	Similarity	1.4%	Score 21	DB 22	Length 2776
Best Local	Similarity	100.0%	Pred. No. 6.3		
Matches	21	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

Oy	1112	AACAGCCCTGGCGCGTGGGC	1132
Db	938	AACAGCCCTGGCGCGTGGGC	918

RESULT 41
AAK68475
ID AAK68475 standard; DNA; 2776 BP.
XX
XX AAK68475;
XX AC
XX
DT 06-NOV-2001 (first entry)
DE
Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23287.
XX
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.

PR	17-MAR-2000	2000US-01090076
PR	18-MAR-2000	2000US-01091233
PR	19-MAR-2000	2000US-01091823
PR	19-MAY-2000	2000US-0205515
PR	07-JUN-2000	2000US-0209647
PR	28-JUN-2000	2000US-02148686
PR	30-JUN-2000	2000US-02151535
PR	07-JUL-2000	2000US-02166347
PR	07-JUL-2000	2000US-02166880
PR	11-JUL-2000	2000US-02174867
PR	11-JUL-2000	2000US-02174966
PR	14-JUL-2000	2000US-0218290
PR	26-JUL-2000	2000US-02209633
PR	26-JUL-2000	2000US-02209664
PR	14-AUG-2000	2000US-02245418
PR	14-AUG-2000	2000US-02245419
PR	14-AUG-2000	2000US-02255113
PR	14-AUG-2000	2000US-02255114
PR	14-AUG-2000	2000US-02255667
PR	14-AUG-2000	2000US-02256567
PR	18-AUG-2000	2000US-02257579
PR	22-AUG-2000	2000US-02266861
PR	22-AUG-2000	2000US-02266868
PR	22-AUG-2000	2000US-022718182
PR	23-AUG-2000	2000US-02270009
PR	30-ANG-2000	2000US-02289824
PR	01-SEP-2000	2000US-02292867
PR	01-SEP-2000	2000US-02293443
PR	01-SEP-2000	2000US-02293444
PR	01-SEP-2000	2000US-02299345
PR	05-SEP-2000	2000US-02299509
PR	06-SEP-2000	2000US-02295133
PR	06-SEP-2000	2000US-02304337
PR	06-SEP-2000	2000US-02304338
PR	08-SEP-2000	2000US-02311442
PR	08-SEP-2000	2000US-02311443
PR	08-SEP-2000	2000US-02311444
PR	08-SEP-2000	2000US-02314113
PR	08-SEP-2000	2000US-02314114
PR	08-SEP-2000	2000US-0232080
PR	08-SEP-2000	2000US-02320801
PR	12-SEP-2000	2000US-02313668
PR	14-SEP-2000	2000US-02323297
PR	14-SEP-2000	2000US-02323298
PR	14-SEP-2000	2000US-02323299
PR	14-SEP-2000	2000US-02324200
PR	14-SEP-2000	2000US-02324201
PR	14-SEP-2000	2000US-02330663
PR	14-SEP-2000	2000US-02330664
PR	14-SEP-2000	2000US-02330665
PR	21-SEP-2000	2000US-02344223
PR	21-SEP-2000	2000US-02344224
PR	21-SEP-2000	2000US-02344274
PR	25-SEP-2000	2000US-0234597
PR	25-SEP-2000	2000US-02345978
PR	26-SEP-2000	2000US-02354848
PR	27-SEP-2000	2000US-02358534
PR	27-SEP-2000	2000US-02358536
PR	29-SEP-2000	2000US-0236327
PR	29-SEP-2000	2000US-02363277
PR	29-SEP-2000	2000US-02363678
PR	29-SEP-2000	2000US-02363679
PR	02-OCT-2000	2000US-02370338
PR	02-OCT-2000	2000US-02370339
PR	02-OCT-2000	2000US-0237040
PR	13-OCT-2000	2000US-02399355

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 08-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0253678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX Disclosure; SEQ ID NO 23287; 3071bp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAM82170 to AAM91921. (I) have cytosstatic

CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK67694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 2776 BP; 545 A; 749 C; 851 G; 631 T; 0 other;
Query Match 1.4%; Score 21; DB 22; Length 2776;
Best Local Similarity 100.0%; Pred. No. 6.3;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1112 AACAGGCCCTGGCGGTGGGC 1132
DB 1839 AACAGGCCCTGGCGGTGGGC 1859
RESULT 42
AAK6476
ID AAK6476 standard; DNA; 2776 BP.
XX
XX AAK6476;
AC
XX 06-NOV-2001 (first entry)
DT
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23288.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cystostatic; gene therapy; vaccine; metastasis; ds.
XX
OS Homo sapiens.
XX
XX WO200157182-A2.
PN
XX 09-AUG-2001.
PD
XX
XX 17-JAN-2001; 2001WO-US01354.
PF
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.

Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1112 AACAGCCCTGGCGGTGGGC 1132
1839 AACAGCCCTGGCGGTGGGC 1859

Db

RESULT 43
ABS69045/C
ID ABS69045 standard; DNA; 352 BP.

XX ABS69045;

XX 21-NOV-2002 (first entry)

XX Novel murine polynucleotide isolated using gene trap technology #108.

XX Mouse; gene trapped sequence; GTS; functional genomic analysis;
XX phage display system; gene chip; temporal gene expression;
XX tissue specific gene expression; antisense inhibition; gene targeting;
XX development disorder; cell differentiation disorder; aging; cancer;
XX autoimmune disease; lupus; inflammatory disorder; skin disorder;
XX degenerative disorder; ds.

XX Mus musculus.

XX US2002102543-A1.

XX 01-AUG-2002.

XX 30-NOV-2000; 2000US-0728445.

XX 01-DEC-1999; 99US-168358P.

XX (FRIE/) FRIEDRICH G.
XX (ZAMB/) ZAMBROWICZ B.
XX (SAND/) SANDS A T.

XX Friedrich G, Zambrowicz B, Sands AT;

XX WPI; 2002-690598/74.

XX Novel murine polynucleotides that individually identify novel genes
XX into which a retroviral gene trap vector has integrated, useful in
XX genomic analysis and in discovery, development of therapeutic and
XX diagnostic agents -

XX Claim 1; Page 57; 296pp; English.

XX The invention describes an isolated murine polynucleotide (i) comprising
XX a contiguous stretch of at least 60 nucleotides of one of 265-677
XX nucleotide 891 OMNIBANK gene trapped sequences (GTS) (S), given in the
XX specification. The novel genes and cells are useful in functional
XX genomic analysis and in the discovery and development of new therapeutic
XX and diagnostic agents and methods. (i) is useful for identifying the
XX coding regions of the murine genome, to isolate cDNAs, genomic clones,
XX or full-length genes/polynucleotides or homologues, heterologues,
XX paralogues, or orthologues that are capable of hybridizing to one or more
XX of the GTS under stringent conditions. (i) can be incorporated into a
XX phage display system that can be used to screen for proteins, or other
XX ligands, that are capable of binding an amino acid sequence encoded by
XX an oligonucleotide or polynucleotide sequence in at least one of the TS
XX sequences. (i) is useful in addressable arrays, such as gene chips, to
XX identify and characterize temporal and tissue specific gene expression,
XX to identify the gene of interest from many sources and for genetic
XX manipulations such as antisense inhibition and gene targeting. Decreasing
XX the level of expression of (i) and/or down regulating the activity of
XX peptides or proteins encoded by (i) is useful for treating development
XX and cell differentiation disorders, aging, cancer, autoimmune disease,
XX lupus, inflammatory disorders, skin disorders and degenerative
XX disorders. This sequence represents a murine cDNA isolated using gene
XX trap technology.

SEQ Sequence 352 BP; 97 A; 89 C; 84 G; 82 T; 0 other;

Query Match

Best local similarity 1.4%; Score 20; DB 24; Length 352;

Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1010 CACAGCCTCATTGGAGAGT 1029
209 CACAGCCTCATTGGAGAGT 190

Db

RESULT 44
ABQ49522/C
ID ABQ49522 standard; DNA; 592 BP.

XX ABQ49522;

XX 12-JUL-2002 (first entry)

XX Oligonucleotide for detecting cytosine methylation SEQ ID NO 36113.

XX Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
XX drug; side effect; cancer; central nervous system; cardiovascular;
XX gastrointestinal; respiratory system; single nucleotide polymorphism;
XX SNP; cell differentiation; ds.

XX Homo sapiens.

XX WO200218632-A2.

XX 07-MAR-2002.

XX 01-SEP-2001; 2001WO-EPI10074.

XX 01-SEP-2000; 2000DE-1043826.

XX 05-SEP-2000; 2000DE-1044543.

XX (EPIG-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K, Guecig D;

XX WPI; 2002-371829/40.

XX Determining the degree of cytosine methylation in genomic DNA, useful
XX for diagnosis and prognosis, comprises selective hybridization of
XX amplicons from chemically treated DNA -

XX Claim 12; 56pp + Sequence Listing; 56pp; German.

XX This invention describes a novel method for determining the degree of
XX methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX genomic sample of DNA. The sample is treated chemically to convert
XX cytosine (C) but not methylated C, to uracil, then part of the genomic
XX DNA that contains the target C is amplified to form a labeled amplicon.
XX The amplicon is hybridized to two classes, each with at least one
XX member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX and the degree of hybridization to both classes is determined from the
XX label on the amplicon. From the ratio of labels hybridized to the two
XX classes of oligomers, the degree of methylation is calculated. The method
XX is used: (i) for diagnosis and/or prognosis of side effects of
XX therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX of the central nervous, cardiovascular, gastrointestinal and respiratory
XX systems etc., particularly by detecting mutations or single nucleotide
XX polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX types and for investigating cell differentiation. The method allows the
XX methylation status of many C residues to be determined simultaneously.
XX AB013410-AB094121 represent genomic DNA sequences used to illustrate the
XX method for determining the degree of cytosine methylation described in
XX the disclosure of the invention.

XX Sequence 592 BP; 81 A; 59 C; 201 G; 251 T; 0 other;

Query Match

1.4%; Score 20; DB 24; Length 592;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 834 CGCTTGGCCCACTACAT 853
DB 68 CGCTTGGCCCACTACAT 49

OY 834 CGCTTGGCCCACTACAT 853
DB 525 CGCTTGGCCCACTACAT 544

Search completed: January 26, 2004, 22:22:39
Job time : 428 secs

RESULT 45

ABQ49523
ID ABQ49523 standard; DNA, 592 BP.

XX AC ABQ49523;

XX DT 12-JUL-2002 (first entry)

XX DE Oligonucleotide for detecting cytosine methylation SEQ ID NO 36114.

XX KW Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;

KW drug; side effect; cancer; central nervous system; cardiovascular;

KW gastrointestinal; respiratory system; single nucleotide polymorphism;

XX SNP; cell differentiation; ds.

XX OS Homo sapiens.

XX PN WO200218632-A2.

XX PD 07-MAR-2002.

XX PF 01-SEP-2001; 2001WO-BP10074.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PR 05-SEP-2000; 2000DE-1044543.

XX PA (EPIC-) EPIGENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K, Guetig D;

XX DR WPI: 2002-371829/40.

XX PT Determining the degree of cytosine methylation in genomic DNA, useful

XX PT for diagnosis and prognosis, comprises selective hybridization of

XX PT amplicons from chemically treated DNA -

XX PS Claim 12; 56pp + Sequence Listing; 56pp; German.

XX CC This invention describes a novel method for determining the degree of
XX CC methylation of a particular cytosine in a motif 5'-CpG-3', present in a
XX CC genomic sample of DNA. The sample is treated chemically to convert
XX CC cytosine (C) but not methylated C, to uracil, then part of the genomic
XX CC DNA that contains the target C is amplified to form a labeled amplicon.
XX CC The amplicon is hybridised to two classes, each with at least one
XX CC member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
XX CC and the degree of hybridisation to both classes is determined from the
XX CC label on the amplicon. From the ratio of labels hybridised to the two
XX CC classes of oligomers, the degree of methylation is calculated. The method
XX CC is used: (i) for diagnosis and/or prognosis of side effects of
XX CC therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
XX CC of the central nervous, cardiovascular, gastrointestinal and respiratory
XX CC systems etc., particularly by detecting mutations or single nucleotide
XX CC polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
XX CC types and for investigating cell differentiation. The method allows the
XX CC methylation status of many C residues to be determined simultaneously.
XX CC ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
XX CC method for determining the degree of cytosine methylation described in
XX CC the disclosure of the invention.

XX SQ Sequence 592 BP; 251 A; 201 C; 59 G; 81 T; 0 other;

Query Match 1.4%; Score 20; DB 24; Length 592;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Result No.	Score	Query Match	Length	DB	ID	Description
1	92	6.3	645	14	US-10-004-717-4	Sequence 4, Appl1
2	92	6.3	861	14	US-10-004-717-24	Sequence 24, Appl1
3	92	6.3	1861	13	US-09-617-360-3	Sequence 3, Appl1
4	36	2.5	65	13	US-09-908-975-1140	Sequence 4140, Appl1
5	32	2.2	65	13	US-09-908-975-30269	Sequence 30269, Appl1
6	32	2.4	5340	9	US-09-617-360-1	Sequence 1, Appl1
7	26	1.8	26	11	US-09-692-665-221	Sequence 221, Appl1
8	23	1.6	738	8	US-08-722-570-13	Sequence 13, Appl1
9	23	1.6	790	14	US-10-004-717-20	Sequence 20, Appl1
10	23	1.6	1385	14	US-10-004-717-30	Sequence 30, Appl1
11	23	1.6	1412	14	US-10-004-717-6	Sequence 6, Appl1
12	23	1.6	1412	14	US-10-004-717-37	Sequence 37, Appl1
13	21	1.4	21	15	US-10-090-011-1	Sequence 1, Appl1
14	21	1.4	31	15	US-10-090-011-19	Sequence 19, Appl1
15	21	1.4	576	13	US-10-020-386-8222	Sequence 8222, Appl1

C	16	21	1.4	723	13	US-10-027-632-129967	Sequence	129767
C	17	21	1.4	723	13	US-10-027-632-129968	Sequence	129768
C	18	21	1.4	723	14	US-10-027-632-129967	Sequence	129767
C	19	21	1.4	723	14	US-10-027-632-129968	Sequence	129768
C	20	21	1.4	2776	11	US-09-764-891-6731	Sequence	6731
C	21	21	1.4	2776	11	US-09-764-891-6733	Sequence	6733
C	22	20	1.4	352	10	US-09-728-445-108	Sequence	108
C	22	20	1.4	352	10	US-09-728-445-108	Sequence	108
C	23	19	1.3	473	11	US-09-918-995-13658	Sequence	13658
C	24	19	1.3	500	10	US-09-783-590-2157	Sequence	2157
C	25	19	1.3	501	12	US-10-260-238-5886	Sequence	5886
C	26	19	1.3	582	13	US-10-027-632-276355	Sequence	276355
C	27	19	1.3	582	14	US-10-027-632-276355	Sequence	276355
C	28	19	1.3	735	13	US-10-027-632-157510	Sequence	157510
C	29	19	1.3	735	14	US-10-027-632-157510	Sequence	157510
C	30	19	1.3	1263	10	US-09-938-8424-1036	Sequence	1036
C	31	19	1.3	1263	12	US-09-938-8424-1036	Sequence	1036
C	32	19	1.3	2180	13	US-10-027-632-111282	Sequence	111282
C	33	19	1.3	2180	13	US-10-027-632-111283	Sequence	111283
C	34	19	1.3	2180	13	US-10-027-632-111284	Sequence	111284
C	35	19	1.3	2180	14	US-10-027-632-111282	Sequence	111282
C	36	19	1.3	2180	14	US-10-027-632-111283	Sequence	111283
C	37	19	1.3	2180	14	US-10-027-632-111284	Sequence	111284
C	38	19	1.3	3424	11	US-09-764-891-8887	Sequence	8887
C	39	19	1.3	3425	11	US-09-764-891-8888	Sequence	8888
C	40	18	1.2	169	12	US-10-242-5356-3008	Sequence	3008
C	41	18	1.2	466	13	US-10-027-632-184196	Sequence	184196
C	42	18	1.2	466	14	US-10-027-632-184196	Sequence	184196
C	43	18	1.2	502	13	US-10-029-386-6219	Sequence	2219
C	44	18	1.2	504	13	US-10-029-386-6504	Sequence	6504
C	45	18	1.2	637	13	US-10-027-632-208140	Sequence	208140

ALIGNMENTS

```

: RESULT 1
: US-10-004-717-4
: Sequence 4, Application US/10004717
: Publication No. US20020192665A1
: GENERAL INFORMATION:
: APPLICANT: ZOGBHI, HUDA Y.
: APPLICANT: YANG, QI
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
: TITLE OF INVENTION: ATONAL ASSOCIATED SEQUENCE FOR DEAFNESS,
: TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
: FILE REFERENCE: P01899US4
: CURRENT APPLICATION NUMBER: US/10/004,717
: CURRENT FILING DATE: 2002-08-16
: PRIOR APPLICATION NUMBER: 09/585,645
: PRIOR FILING DATE: 2000-06-01
: PRIOR APPLICATION NUMBER: 60/176,993
: PRIOR FILING DATE: 2000-01-19
: PRIOR APPLICATION NUMBER: 60/137,060
: PRIOR FILING DATE: 1999-06-01
: NUMBER OF SEQ ID NOS: 69
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 4
: LENGTH: 645
: TYPE: DNA
: ORGANISM: Mus musculus
: US-10-004-717-4

```

Query Match	6.3%;	Score 92;	DB 14;	Length 645;
Best Local Similarity	100.0%;	Pred. No. 3.9e-37;		
Matches	92;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Oy	762	GCGCTGGATGCGCCTTCCGCAGTGTCCTGCCCAACTTCCGGATGACGCAAACTTTACAAG	821
Db	304	GCGCTGGATGCGCCTTCCGCAGTGTCCTGCCCAACTTCCGGATGACGCAAACTTTACAAG	363
Oy	822	ATTGGAACCTTGGCTTTGGCCCCAACACTACTAT	853
Db	364	ATTGGAACCTTGGCTTTGGCCCCAACACTACTAT	395

RESULT 2

US-10-004-717-24
; Sequence 24, Application US/10004717
; Publication No. US2002019265A1
; GENERAL INFORMATION:
; APPLICANT: ZOGBHI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899US4
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 24
; LENGTH: 861
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-24

Query Match 6.3%; Score 92; DB 14; Length 861;
Best Local Similarity 100.0%; Pred. No. 3.8e-37;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GCGCTGATGCGCTGCGCGGTGTCTGCGCCACCTTCCCGAGAGACCCAACTTACAAAG 821
DB 463 GCGCTGATGCGCTGCGCGGTGTCTGCGCCACCTTCCCGAGAGACCCAACTTACAAAG 522
OY 822 ATCGAGACCTTGCGCTTCCGCCACCACTACAT 853
DB 523 ATCGAGACCTTGCGCTTCCGCCACCACTACAT 554

RESULT 3

US-09-817-360-3
; Sequence 3, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; APPLICANT: Lin, Joseph
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; CURRENT FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1861
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-817-360-3

Query Match 6.3%; Score 92; DB 9; Length 1861;
Best Local Similarity 100.0%; Pred. No. 3.5e-37;

Matches 92; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 762 GCGCTGATGCGCTGCGCGGTGTCTGCGCCACCTTCCCGAGAGACCCAACTTACAAAG 821
DB 1396 GCGCTGATGCGCTGCGCGGTGTCTGCGCCACCTTCCCGAGAGACCCAACTTACAAAG 1455

OY 822 ATCGAGACCTTGCGCTTCCGCCACCACTACAT 853

DB 1456 ATCGAGACCTTGCGCTTCCGCCACCACTACAT 1487

RESULT 4

US-09-908-975-4140
; Sequence 4140, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, AVI
; APPLICANT: WASSERMAN, ALON
; APPLICANT: MINTZ, ELI
; APPLICANT: MINTZ, LIAT
; APPLICANT: FAIGLER, SIMCHON
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4140
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Rattus norvegicus
US-09-908-975-4140

Query Match 2.5%; Score 36; DB 13; Length 65;
Best Local Similarity 100.0%; Pred. No. 5.9e-08;

Matches 36; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1281 ATTGAGGCTGATCTCTTAAACCTCTCAGTGT 1316
DB 30 ATTGAGGCTGATCTCTTAAACCTCTCAGTGT 65

RESULT 5

US-09-908-975-30269
; Sequence 30269, Application US/09908975
; Publication No. US20030165843A1
; GENERAL INFORMATION:
; APPLICANT: SHOSHAN, AVI
; APPLICANT: WASSERMAN, ALON
; APPLICANT: MINTZ, ELI
; APPLICANT: MINTZ, LIAT
; APPLICANT: FAIGLER, SIMCHON
; TITLE OF INVENTION: OLIGONUCLEOTIDE LIBRARY FOR DETECTING RNA TRANSCRIPTS AND SPLIC
; TITLE OF INVENTION: THAT POPULATE A TRANSCRIPTOME
; FILE REFERENCE: 36688-0005
; CURRENT APPLICATION NUMBER: US/09/908,975
; CURRENT FILING DATE: 2001-07-20
; PRIOR APPLICATION NUMBER: US 60/287,724
; PRIOR FILING DATE: 2001-05-02
; PRIOR APPLICATION NUMBER: US 60/221,607
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 32337
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30269
; LENGTH: 65
; TYPE: DNA
; ORGANISM: Mus musculus
US-09-908-975-30269

Query Match 2.4%; Score 35; DB 13; Length 65;
Best Local Similarity 100.0%; Pred. No. 1.9e-07;

Matches 35; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 547 TCAGTTCATTCACCCGACCTAGCCCACTCTC 581
Db 1 TCAGTTCATTCACCCGACCTAGCCCACTCTC 35

RESULT 6

US-09-817-360-1
; Sequence 1, Application US/09817360
; Patent No. US20020015696A1
; GENERAL INFORMATION:
; APPLICANT: German, Michael S.
; TITLE OF INVENTION: PRODUCTION OF PANCREATIC ISLET CELLS
; TITLE OF INVENTION: AND DELIVERY OF INSULIN
; FILE REFERENCE: UCSF-129CIP
; CURRENT APPLICATION NUMBER: US/09/817,360
; PRIOR FILING DATE: 2001-03-20
; PRIOR APPLICATION NUMBER: 09/535,145
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: 60/128,180
; PRIOR FILING DATE: 1999-04-06
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 5340
; TYPE: DNA
; ORGANISM: Homo Sapiens
US-09-817-360-1

Query Match 2.2%; Score 32; DB 9; Length 5340;
Best Local Similarity 100.0%; Pred. No. 4.5e-06;

Matches 32; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 672 AAGAGCGAGTTGGCAGCTGAGCAGAGCGAGC 703
Db 3235 AAGAGCGAGTTGGCAGCTGAGCAGAGCGAGC 3266

RESULT 7

US-09-992-665-221/c
; Sequence 221, Application US/09992665
; Publication No. US20030092009A1
; GENERAL INFORMATION:
; APPLICANT: Kaia Palm
; TITLE OF INVENTION: PROFILING TUMOR SPECIFIC MARKERS FOR THE
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF NEOPLASTIC DISEASE
; FILE REFERENCE: CEMINIS.002A
; CURRENT APPLICATION NUMBER: US/09/992,665
; CURRENT FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: 60/249,508
; PRIOR FILING DATE: 2000-11-16
; NUMBER OF SEQ ID NOS: 380
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 221
; LENGTH: 26
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Probe
US-09-992-665-221

Query Match 1.8%; Score 26; DB 11; Length 26;
Best Local Similarity 100.0%; Pred. No. 0.01;
Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 673 AGAGCGAGTTGGCAGCTGAGCAGAGCAG 698
Db 26 AGAGCGAGTTGGCAGCTGAGCAGAGCAG 1

RESULT 8

US-08-722-570-13
; Sequence 13, Application US/08722570

; Publication No. US20030044887A1

; GENERAL INFORMATION:

; APPLICANT: Anderson, David J.

; APPLICANT: Ma, Qifu

; TITLE OF INVENTION: NEUROGENIN

; NUMBER OF SEQUENCES: 20

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Flehr, Hobbach, Teat, Albritton & Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/722,570

; FILING DATE: 27-SEP-1996

; CLASSIFICATION: 5365

; ATTORNEY/AGENT INFORMATION:

; NAME: Silva, Robin M.

; REGISTRATION NUMBER: 38,304

; REFERENCE/DOCKET NUMBER: A-63902/RFT/RMS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELETYPE: 910 277299

; INFORMATION FOR SEQ ID NO: 13:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 738 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: DNA

US-08-722-570-13

Query Match 1.6%; Score 23; DB 8; Length 738;
Best Local Similarity 100.0%; Pred. No. 0.26;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 732 GAGCGCAACCGCATGACCAACT 754
Db 304 GAGCGCAACCGCATGACCAACT 326

RESULT 9

US-10-004-717-20
; Sequence 20, Application US/10004717
; Publication No. US20020192665A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, QI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEPRESS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 20
; LENGTH: 790
; TYPE: DNA
; ORGANISM: chicken

US-10-004-717-20

Query Match 1.6%; Score 23; DB 14; Length 790;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
DB 374 GAGCGCAACCGCATGCACACCT 396

RESULT 10
US-10-004-717-30

; Sequence 30, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 30
; LENGTH: 1385
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-30

Query Match 1.6%; Score 23; DB 14; Length 1385;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
DB 742 GAGCGCAACCGCATGCACACCT 764

RESULT 11
US-10-004-717-6

; Sequence 6, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, OI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-6

Query Match 1.6%; Score 23; DB 14; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
DB 445 GAGCGCAACCGCATGCACACCT 467

RESULT 12
US-10-004-717-37

; Sequence 37, Application US/10004717
; Publication No. US20020192655A1
; GENERAL INFORMATION:
; APPLICANT: ZOGHBI, HUDA Y.
; APPLICANT: YANG, OI
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPEUTIC USE OF AN
; TITLE OF INVENTION: ATOMAL ASSOCIATED SEQUENCE FOR DEAFNESS,
; TITLE OF INVENTION: OSTEOARTHRITIS, AND ABNORMAL CELL PROLIFERATION
; FILE REFERENCE: P01899054
; CURRENT APPLICATION NUMBER: US/10/004,717
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 09/585,645
; PRIOR FILING DATE: 2000-06-01
; PRIOR APPLICATION NUMBER: 60/176,993
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: 60/137,060
; PRIOR FILING DATE: 1999-06-01
; NUMBER OF SEQ ID NOS: 69
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 37
; LENGTH: 1412
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-004-717-37

Query Match 1.6%; Score 23; DB 14; Length 1412;
Best Local Similarity 100.0%; Pred. No. 0.24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACACCT 754
DB 445 GAGCGCAACCGCATGCACACCT 467

RESULT 13
US-10-090-011-1

; Sequence 1, Application US/10090011
; Publication No. US20030082810A1
; GENERAL INFORMATION:
; APPLICANT: Serup, Palle
; APPLICANT: Heimberg, Harry
; APPLICANT: Gradwohl, Gerard
; TITLE OF INVENTION: Methods For Generating Insulin-Secreting
; TITLE OF INVENTION: Cells Suitable for Transplantation
; FILE REFERENCE: 6246.200-US
; CURRENT APPLICATION NUMBER: US/10/090,011
; CURRENT FILING DATE: 2002-02-26
; PRIOR APPLICATION NUMBER: US 60/271,474
; PRIOR FILING DATE: 2001-02-26
; NUMBER OF SEQ ID NOS: 70
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 21
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-090-011-1

Query Match 1.4%; Score 21; DB 15; Length 21;
Best Local Similarity 100.0%; Pred. No. 4.1;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 460 TGGGCGCTCATCCCTTGATG 480
 DB 1 TGGGCGCTCATCCCTTGATG 21

DB 29 AACAGGCGCTTGCGCGGTGGGC 49
 Search completed: January 29, 2004, 22:30:14
 Job time : 568 secs

RESULT 14

US-10-090-011-49
 ; Sequence 49, Application US/10090011
 ; Publication No. US20030082810A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Serup, Palle
 ; APPLICANT: Heimberg, Harry
 ; APPLICANT: Gradwohl, Gerard
 ; TITLE OF INVENTION: Methods For Generating Insulin-Secreting
 ; TITLE OF INVENTION: Cells Suitable for Transplantation
 ; FILE REFERENCE: 6246.200-US
 ; CURRENT APPLICATION NUMBER: US/10/090.011
 ; CURRENT FILING DATE: 2002-02-26
 ; PRIOR APPLICATION NUMBER: US 60/271,474
 ; PRIOR FILING DATE: 2001-02-26
 ; NUMBER OF SEQ ID NOS: 70
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 49
 ; LENGTH: 21
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-090-011-49

Query Match 1.4%; Score 21; DB 15; Length 21;
 Best Local Similarity 100.0%; Pred. No. 4.1;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 798 CCGGATGACGCCAATTACA 818
 DB 1 CCGGATGACGCCAATTACA 21

RESULT 15

US-10-029-386-8222
 ; Sequence 8222, Application US/10029386
 ; Publication No. US20030194704A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Penn, Sharon G.
 ; APPLICANT: Rank, David R.
 ; APPLICANT: Hanzel, David K.
 ; TITLE OF INVENTION: HUMAN GENOME-DRIVEN SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
 ; FILE REFERENCE: AEOMICA-X-2
 ; CURRENT APPLICATION NUMBER: US/10/029.386
 ; CURRENT FILING DATE: 2001-12-20
 ; NUMBER OF SEQ ID NOS: 34288
 ; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
 ; SEQ ID NO 8222
 ; LENGTH: 576
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: MAP TO CHR16.3
 ; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 67
 ; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 64
 ; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 9.8
 ; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 41
 ; OTHER INFORMATION: EST_HUMAN HIT: BG166801.1, EVALU0.00e+00
 ; OTHER INFORMATION: NT HIT: g14779902, EVALU0.00e-95
 ; OTHER INFORMATION: SWISSPROT HIT: P25444, EVALU0.00e-58
 US-10-029-386-8222

Query Match 1.4%; Score 21; DB 13; Length 576;
 Best Local Similarity 100.0%; Pred. No. 2.9;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1112 AACAGGCGCTTGCGCGGTGGGC 1132
 |||||

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004. Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 29, 2004, 19:42:42 ; Search time 2899 Seconds
(without alignments)
12240.267 Million cell updates/sec

Title: US-09-595-947E-1

Perfect score: 1460
Sequence: 1 gcacgtacgcagagagagcagc.....agagtcacccatccagctgt 1460

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrt1:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	92	6.3	593	28	AZ296526
2	60	4.1	1025	14	BY708009
3	60	4.1	1540	11	AK008017
4	24	1.6	923	13	BX445903

5	23	1.6	504	4	BX524720	BX524720 RZPD Mus
6	23	1.6 <td>517</td> <td>13 <td>BU053975</td> <td>BU053975 UI-M-PD0-</td> </td>	517	13 <td>BU053975</td> <td>BU053975 UI-M-PD0-</td>	BU053975	BU053975 UI-M-PD0-
7	23	1.6 <td>600</td> <td>12 <td>BG808248</td> <td>BG808248 2083-52 M</td> </td>	600	12 <td>BG808248</td> <td>BG808248 2083-52 M</td>	BG808248	BG808248 2083-52 M
8	23	1.6 <td>600</td> <td>13 <td>BU924937</td> <td>BU924937 7103-91 M</td> </td>	600	13 <td>BU924937</td> <td>BU924937 7103-91 M</td>	BU924937	BU924937 7103-91 M
9	23	1.6 <td>673</td> <td>13 <td>BU709606</td> <td>BU709606 UI-M-FR0-</td> </td>	673	13 <td>BU709606</td> <td>BU709606 UI-M-FR0-</td>	BU709606	BU709606 UI-M-FR0-
10	23	1.6 <td>687</td> <td>14 <td>CA945402</td> <td>CA945402 UI-M-PD0-</td> </td>	687	14 <td>CA945402</td> <td>CA945402 UI-M-PD0-</td>	CA945402	CA945402 UI-M-PD0-
11	23	1.6 <td>689</td> <td>13 <td>BU058877</td> <td>BU058877 UI-M-FR0-</td> </td>	689	13 <td>BU058877</td> <td>BU058877 UI-M-FR0-</td>	BU058877	BU058877 UI-M-FR0-
12	23	1.6 <td>730</td> <td>13 <td>BU612495</td> <td>BU612495 UI-M-FR0-</td> </td>	730	13 <td>BU612495</td> <td>BU612495 UI-M-FR0-</td>	BU612495	BU612495 UI-M-FR0-
13	23	1.6 <td>814</td> <td>13 <td>BQ178789</td> <td>BQ178789 UI-M-EV0-</td> </td>	814	13 <td>BQ178789</td> <td>BQ178789 UI-M-EV0-</td>	BQ178789	BQ178789 UI-M-EV0-
14	23	1.6 <td>823</td> <td>13 <td>BU054481</td> <td>BU054481 UI-M-PD0-</td> </td>	823	13 <td>BU054481</td> <td>BU054481 UI-M-PD0-</td>	BU054481	BU054481 UI-M-PD0-
15	23	1.6 <td>932</td> <td>14 <td>CA979119</td> <td>CA979119 AGENCOURT</td> </td>	932	14 <td>CA979119</td> <td>CA979119 AGENCOURT</td>	CA979119	CA979119 AGENCOURT
16	23	1.6 <td>1001</td> <td>9 <td>AL540071</td> <td>AL540071 AL540071</td> </td>	1001	9 <td>AL540071</td> <td>AL540071 AL540071</td>	AL540071	AL540071 AL540071
17	23	1.6 <td>1269</td> <td>12 <td>BG854922</td> <td>BG854922 1024041CO</td> </td>	1269	12 <td>BG854922</td> <td>BG854922 1024041CO</td>	BG854922	BG854922 1024041CO
18	22	1.5 <td>589</td> <td>13 <td>BU775885</td> <td>BU775885 S.EBRD03</td> </td>	589	13 <td>BU775885</td> <td>BU775885 S.EBRD03</td>	BU775885	BU775885 S.EBRD03
19	21	1.4	319	10 <td>BE936551</td> <td>BE936551 RCS-NT005</td>	BE936551	BE936551 RCS-NT005
20	21	1.4	464	28 <td>AQ753599</td> <td>AQ753599 HS 2117 A</td>	AQ753599	AQ753599 HS 2117 A
21	21	1.4	536	28 <td>AZ506899</td> <td>AZ506899 1M0348G13</td>	AZ506899	AZ506899 1M0348G13
22	21	1.4	704	13 <td>BU057851</td> <td>BU057851 UI-M-FR0-</td>	BU057851	BU057851 UI-M-FR0-
23	21	1.4	718	10 <td>BE263765</td> <td>BE263765 601194122</td>	BE263765	BE263765 601194122
24	21	1.4	740	14 <td>CA319439</td> <td>CA319439 UI-M-FW0-</td>	CA319439	CA319439 UI-M-FW0-
25	21	1.4	781	13 <td>BU611678</td> <td>BU611678 UI-M-FR0-</td>	BU611678	BU611678 UI-M-FR0-
26	21	1.4	829	12 <td>BI910102</td> <td>BI910102 603067946</td>	BI910102	BI910102 603067946
27	21	1.4	835	14 <td>CA320553</td> <td>CA320553 UI-M-FW0-</td>	CA320553	CA320553 UI-M-FW0-
28	21	1.4	913	10 <td>BF204175</td> <td>BF204175 601867625</td>	BF204175	BF204175 601867625
29	21	1.4	953	9 <td>AU067624</td> <td>AU067624 AU067624</td>	AU067624	AU067624 AU067624
30	21	1.4	962	13 <td>BQ686909</td> <td>BQ686909 AGENCOURT</td>	BQ686909	BQ686909 AGENCOURT
31	21	1.4	1022	12 <td>BW563662</td> <td>BW563662 AGENCOURT</td>	BW563662	BW563662 AGENCOURT
32	21	1.4	1039	10 <td>BE780690</td> <td>BE780690 601469349</td>	BE780690	BE780690 601469349
33	21	1.4	1042	10 <td>BG419220</td> <td>BG419220 602445870</td>	BG419220	BG419220 602445870
34	20	1.4	241	9 <td>AA113743</td> <td>AA113743 WSE Pyroc</td>	AA113743	AA113743 WSE Pyroc
35	20	1.4	351	14 <td>T70947</td> <td>T70947 yc49c05.r1</td>	T70947	T70947 yc49c05.r1
36	20	1.4	397	12 <td>BI445873</td> <td>BI445873 da133c04</td>	BI445873	BI445873 da133c04
37	20	1.4	446	9 <td>AV944801</td> <td>AV944801 AV944801</td>	AV944801	AV944801 AV944801
38	20	1.4	476	12 <td>BM253695</td> <td>BM253695 514849 MA</td>	BM253695	BM253695 514849 MA
39	20	1.4	550	9 <td>AV939670</td> <td>AV939670 AV939670</td>	AV939670	AV939670 AV939670
40	20	1.4	909	14 <td>CD325295</td> <td>CD325295 AGENCOURT</td>	CD325295	CD325295 AGENCOURT
41	20	1.4	971	29 <td>CNS01YSD</td> <td>AL173398 Tetraodon</td>	CNS01YSD	AL173398 Tetraodon
42	20	1.4	971	29 <td>CNS02YDS</td> <td>AL219506 Tetraodon</td>	CNS02YDS	AL219506 Tetraodon
43	20	1.4	1010	13 <td>BX377288</td> <td>BX377288 BX377288</td>	BX377288	BX377288 BX377288
44	20	1.4	1075	12 <td>BM810917</td> <td>BM810917 AGENCOURT</td>	BM810917	BM810917 AGENCOURT
45	19	1.3	201	10 <td>BB071548</td> <td>BB071548 BB071548</td>	BB071548	BB071548 BB071548

ALIGNMENTS

RESULT 1
AZ296526/c
LOCUS
DEFINITION
PCIC-23-160G18.TV PCIC-23 Mus musculus genomic clone PCIC-23-160G18
' genomic survey sequence.
ACCESSION
AZ296526
VERSION
AZ296526.1 GI:9538311
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 593)
REFERENCE
Zhao, S., Niernan, W., Feldblum, T., Malek, J., Shatman, S., Akintet
B., Levine, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C. M.
Mouse BAC End Sequences from Library PCIC-23
Unpublished
Other GSSs: PCIC-23-160G18.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhaoc@igf.org
Clones are derived from the mouse BAC library PCIC-23. For BAC

JOURNAL COMMENT
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/INLNL at: <http://image.lnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project (BMAP)
 Seq primer: pYX-5.
FEATURES
 source
 1..517
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6403258"
 /tissue_type="whole brain"
 /dev_stage="embryo 12.5 dpc"
 /lab_host="MDA10B (T1 phage resistant)"
 /clone_lib="NIH BMAP FDO"
 /note="Organ: brain; Vector: pYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured mRNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP). Gene Discovery in the Developing Mouse Nervous System", supported by National Institutes of Mental Health (NIMH), Hemin Chin, Ph.D., program coordinator."

BASE COUNT
 ORIGIN
 90 a 166 c 194 g 67 t

Query Match
 Best Local Similarity 100.0%; Score 23; DB 13; Length 517;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
 732 GAGCGCAACCGCATGCACAACCT 754
 |||||
 449 GAGCGCAACCGCATGCACAACCT 471

RESULT 7
LOCUS
 BG808248 600 bp mRNA linear EST 20-DEC-2001
DEFINITION
 2083-52 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,
 mRNA sequence.
ACCESSION
 BG808248
VERSION
 BG808248.1 GI:17955225
KEYWORDS
 EST.
SOURCE
 Mus musculus (house mouse)
ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
REFERENCE
 1 (bases 1 to 600)
 Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
 Gene expression in the developing mouse retina by EST sequencing
 and microarray analysis
 Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL MEDLINE PUBMED COMMENT
 21671825
 11812828
 Contact: Klein WH
 Department of Biochemistry and Molecular Biology

FEATURES
 source
 1..600
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="neural retina"
 /dev_stage="embryonic day 14.5 post-fertilization"
 /clone_lib="Mouse E14.5 retina lambda ZAP II library"

BASE COUNT
 ORIGIN
 99 a 194 c 222 g 82 t 3 others

Query Match
 Best Local Similarity 100.0%; Score 23; DB 13; Length 600;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
 732 GAGCGCAACCGCATGCACAACCT 754
 |||||
 363 GAGCGCAACCGCATGCACAACCT 385

RESULT 9
LOCUS
 BU709606 673 bp mRNA linear EST 26-NOV-2002

FEATURES
 source
 1..600
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 /tissue_type="neural retina"
 /dev_stage="embryonic day 14.5 post-fertilization"
 /clone_lib="Mouse E14.5 retina lambda ZAP II library"

BASE COUNT
 ORIGIN
 94 a 238 c 161 g 107 t

Query Match
 Best Local Similarity 100.0%; Score 23; DB 12; Length 600;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db
 732 GAGCGCAACCGCATGCACAACCT 754
 |||||
 4 GAGCGCAACCGCATGCACAACCT 26

RESULT 8
LOCUS
 BU924937 600 bp mRNA linear EST 30-OCT-2002
DEFINITION
 7103-91 Mouse E14.5 retina lambda ZAP II library Mus musculus cDNA,
 mRNA sequence.
ACCESSION
 BU924937
VERSION
 BU924937.1 GI:24428820
KEYWORDS
 EST.
SOURCE
 Mus musculus (house mouse)
ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murine; Mus.
 1 (bases 1 to 600)
 Mu, X., Zhao, S., Pershad, R., Hsieh, T.-F., Scarpa, A., Wang, S.W.,
 White, R.A., Beremand, P.D., Thomas, T.L., Gan, L. and Klein, W.H.
 Gene expression in the developing mouse retina by EST sequencing
 and microarray analysis
 Nucleic Acids Res. 29 (24), 4983-4993 (2001)

JOURNAL MEDLINE PUBMED COMMENT
 21671825
 11812828
 Contact: Klein WH
 Department of Biochemistry and Molecular Biology
 University of Texas M.D. Anderson Cancer Center
 Box 117, 1515 Holcombe Blvd., Houston, TX 77030, USA
 Tel: 713 792 3646
 Fax: 713 790 0329.

DEFINITION UI-M-FRO-cbe-9-01-0-UI.r1 NIH BMAP_FRO Mus musculus cDNA clone
IMAGE: 6808442 5', mRNA sequence.

ACCESSION BU709606
VERSION BU709606.1 GI:23643245
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 673)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. Jim Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source
Seq primer: PYX-5.
Location/Qualifiers
1..673
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6808442"
/tissue_type="whole brain"
/dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
/lab_host="MDH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FRO"
/note="Organ: Brain; Vector: PYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACG. This library was created for the University of Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemn Chin, Ph.D., program coordinator."

BASE COUNT 122 a 219 c 250 g 81 t 1 others
ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 673;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754
|||||
Db 570 GAGCGCAACCGCATGCACAACT 592
|||||

RESULT 10
LOCUS CA945402 687 bp mRNA linear EST 30-DEC-2002
DEFINITION UI-M-FRO-cdh-1-12-0-UI.r1 NIH BMAP_FRO Mus musculus cDNA clone
IMAGE: 6828925 5', mRNA sequence.

ACCESSION CA945402
VERSION CA945402.1 GI:27433882
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 687)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Dr. James Lin, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
This clone was contributed by the Brain Molecular Anatomy Project (BMAP)

FEATURES
source
Seq primer: PYX-5.
Location/Qualifiers
1..687
/organism="Mus musculus"
/mol_type="mRNA"
/strain="CS7BL/6"
/db_xref="taxon:10090"
/clone="IMAGE: 6828925"
/tissue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="MDH10B (T1 phage resistant)"
/clone_lib="NIH BMAP_FRO"
/note="Organ: brain; Vector: PYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with Not I, and then cloned directionally into PYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is TGAGAGAGCC. This library was created for the University of Iowa Mouse Brain Molecular Anatomy Project (BMAP): 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institutes of Mental Health (NIMH), Hemn Chin, Ph.D., program coordinator."

BASE COUNT 117 a 226 c 250 g 91 t 1 others
ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 687;
Best Local Similarity 100.0%; Pred. No. 5;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 732 GAGCGCAACCGCATGCACAACT 754
|||||
Db 449 GAGCGCAACCGCATGCACAACT 471
|||||

RESULT 11
LOCUS BU058877 689 bp mRNA linear EST 26-AUG-2002
DEFINITION UI-M-FRO-cak-k-23-0-UI.r1 NIH BMAP_FRO Mus musculus cDNA clone
IMAGE: 6413710 5', mRNA sequence.

ACCESSION BU058877
VERSION BU058877.1 GI:22499166
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 689)
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.
TITLE NIH-MGC
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished
Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES

source

Seq primer: PYX-5.

Location/Qualifiers

1. .689
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="IMAGE:6413710"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_id="NTH_BMAP_FR0"
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into PYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 128 a 228 c 255 g 78 t
 ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 689;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACT 754
 Db 626 GAGCGCAACCGCATGCACAACT 648

RESULT 12

LOCUS BU612495 730 bp mRNA linear EST 20-FEB-2003
 DEFINITION UI-M-FR0-cbc-k-21-0-UI.r1 NIH_BMAP_FR0 Mus musculus cDNA clone
 BU612495

ACCESSION BU612495.1 GI:23278710
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 730)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. Jim Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be obtained
 from Dr. M. Bento Soares, bento-soares@iowa.edu

FEATURES

source

This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)
 Seq primer: PYX-5.

Location/Qualifiers

1. .730
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6"
 /db_xref="taxon:10090"
 /clone="UI-M-FR0-cbc-k-21-0-UI"
 /tissue_type="whole brain"
 /dev_stage="embryo 13.5,14.5,16.5,17.5dpc"
 /lab_host="DH10B (T1 phage resistant)"
 /clone_id="NTH_BMAP_FR0"
 /note="Organ: Brain; Vector: PYX-Asc; Site 1: EcoR I;
 Site 2: Not I; The library was constructed according
 Bonaldi, Lennon and Soares, Genome Research, 6:791-806,
 1996. Denatured RNA was size fractionated on a 1% agarose
 gel. First strand cDNA synthesis was primed with oligo-dT
 primer containing a Not I site. Double strand cDNA was
 size selected according to mRNA size fraction, ligated
 with EcoR I adaptor, digested with NotI and then cloned
 directionally into PYX-Asc vector. The library tag
 sequence located between the Not I site and the polyA tail
 is AGCGAGACAG. This library was created for the University
 Iowa Brain Anatomy Project (BMAP): 'Gene Discovery in the
 Developing Mouse Nervous System', supported by National
 Institute of Mental Health (NIMH), Hemin Chin, Ph.D.,
 program coordinator."

BASE COUNT 131 a 244 c 264 g 91 t
 ORIGIN

Query Match 1.6%; Score 23; DB 13; Length 730;
 Best Local Similarity 100.0%; Pred. No. 5;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 732 GAGCGCAACCGCATGCACAACT 754
 Db 570 GAGCGCAACCGCATGCACAACT 592

RESULT 13

LOCUS BQ178789 814 bp mRNA linear EST 30-APR-2002
 DEFINITION UI-M-EV0-bwc-k-07-0-UI.r1 NIH_BMAP_EV0 Mus musculus cDNA clone
 IMAGE:5701758 5', mRNA sequence.

ACCESSION BQ178789
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

Mus musculus (house mouse)
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 814)
 NIH-MGC <http://mgi.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Dr. James Lin, University of Iowa
 cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 This clone was contributed by the Brain Molecular Anatomy Project
 (BMAP)

FEATURES

source

Seq primer: PYX-5.
 Location/Qualifiers
 1. .814
 /organism="Mus musculus"
 /mol_type="mRNA"

```
location/Qualifiers
1. 823
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/cid="IMAGE:640447"
/risue_type="whole brain"
/dev_stage="embryo 12.5 dpc"
/lab_host="Dh10B (T1 phage resistant)"
clone_id="W11_BM4_F0"
```

/clone_11b="NIH MGC 164"
 /notes="Vector: pCMV-Sport6.1.ccd; Site 1: EcoRV; Site 2:
 NotI; Non-normalized, full-length enriched library from
 pooled mouse embryonic limb, maxilla and mandible, day
 10.5 and 11.5 (size selected for the 0.5-1 kb fragments)
 Cloned directionally, priming method: Oligo-dr. cDNA
 enrichment: >1k bp, Average insert size 1.8k bp. Priming
 sequence: 5'-GACTAGTCTTAGATCCGAGCGAGCCCTT-3'. Tissue
 contributed by, David Rowe. Library constructed by ResGen
 Invitrogen Corp."

ORIGIN

Query Match 1.6%; Score 23; DB 14; Length 932;
 Best Local Similarity 100.0%; Pred. No. 5.2;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 732 GAGCGCAACCGCATGCACACCT 754
 |||||
 Db 561 GAGCGCAACCGCATGCACACCT 583

Search completed: January 29, 2004, 22:18:31
 Job time : 2908 secs

THIS PAGE BLANK (USPTO)